THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 38 | May, 1947 | Number 5

GROWTH IN THE GENERAL AND VERBAL BI-FACTORS FROM GRADE VII TO GRADE IX

FRANCES SWINEFORD

The University of Chicago

The bi-factor method of factor analysis has now been developed to the point where the factor composition of a number of mental tests is known and the stability of a solution has been demonstrated. Until recently, however, virtually no facts have been available regarding the growth of the several factors. A cross-section type of study dealing with growth of the general factor and the verbal and spatial group factors measured by a battery of nine tests has been completed.[1] The advantages that would accrue were data available for the same individuals from year to year were recognized at the time the study was undertaken.

The writer is now in possession of a second set of scores on six of the tests for pupils from two of the original grade-groups. Group A consists of two hundred twelve pupils who were tested in the eighth grades of seven different schools in the Autumn of 1944 and a year later in the ninth grade in the high school which serves the same community. Group B includes one hundred seventy-three pupils in the seventh grade in 1944 and in the ninth grade in 1946.

The tests, which can be administered in two or three sittings totalling about two hours' time, include three which measure the general intellective bi-factor only; namely, (1) Arithmetic, (2) Series Completion, and (3) Deduction, and three which measure both the general and the verbal bi-factors; namely, (4) Gen-

[1] Frances Swineford, *The Nature of the General, Verbal, and Spatial Bi-factors.* Unpublished Doctor's thesis, Department of Education, University of Chicago, 1946.

eral Information, (5) Reading Comprehension, and (6) Word Meaning.

With these data it is possible to investigate with greater confidence some of the relationships and trends indicated in the cross-section study. First, however, there must be settled the question of possible practice effects due to the repeated use of the same testing instruments.

Since the two testings for Group A were separated by a year's time, the increases in scores may be due to practice or to mental growth—probably to both. The best indication of practice effects, therefore, is the difference between the ninth-grade means for pupils taking the tests for the very first time and the means for those being tested after the one-year interval. Of the eight hundred four pupils tested in Grade IX in 1945, five hundred eighty-one were taking the tests for the first time, eleven had taken some of the tests, but not all, in 1944, and the remainder, of course, are Group A. For each test the mean score for those taking it for the first time and the mean score for those who had taken it the previous year were recorded by sex, the differences found, and the critical ratios computed. The results are presented in Table 1.

With the exception of Test 1, all mean differences favor the group who had tried the tests the previous year. The differences, however, are not large, although the critical ratios show most of them to be too great to be attributable to chance alone. That they are not the consequence of some sort of selection related to either the general or the verbal factor is indicated by the fact that no significant differences appear for Test 1, which affords an excellent measure of the general factor, nor for Test 6, which provides the best measure of the verbal factor. On the other hand, these two tests are similar to the kinds of tasks regularly required of pupils. The remaining tests are more likely to be quite new to the majority and might, therefore, necessitate a few moments of 'warming up' on the part of those who had never before seen them.

Similar data on the first three tests have been prepared for the 1946 ninth-grade groups. Mean scores for five hundred eighty pupils taking these tests for the first time and one hundred ninety-three pupils taking them for the second time have been computed. The results, given in Table 2, support the foregoing

TABLE 1.—COMPARISONS OF NINTH-GRADE TEST SCORES FOR PUPILS WHO WERE TESTED FOR THE FIRST TIME IN 1945 AND PUPILS WHO HAD BEEN TESTED IN 1944 ALSO

Test	Mean Score		Difference between Means	Standard Error of Difference	Critical Ratio
	Group Tested in 1945 Only	Group Tested in 1944 and 1945			
1 Arithmetic:					
Girls	16.93	17.50	0.57	0.620	0.92
Boys	16.09	15.24	− .85	.538	−1.58
All cases	16.50	16.31	− .19	.416	− .46
2 Series Completion:					
Girls	14.28	16.99	2.71	1.079	2.51
Boys	16.26	18.68	2.42	1.169	2.07
All cases	15.29	17.87	2.58	.802	3.22
3 Deduction:					
Girls	30.67	33.12	2.45	.735	3.33
Boys	31.64	33.85	2.21	.691	3.20
All cases	31.16	33.50	2.34	.504	4.64
4 General Information:					
Girls	31.54	33.32	1.78	.854	2.08
Boys	33.16	35.06	1.90	.893	2.13
All cases	32.36	34.26	1.90	.624	3.04
5 Reading Comprehension:					
Girls	13.08	14.23	1.15	.614	1.87
Boys	12.87	14.21	1.34	.518	2.59
All cases	12.97	14.22	1.15	.398	3.14
6 Word Meaning:					
Girls	21.70	22.70	1.00	.942	1.06
Boys	21.99	22.32	.33	1.017	.32
All cases	21.85	22.49	0.64	0.699	0.92

theory. Test 1 again provides no evidence at all of practice effects. The differences for Test 2 appear to be no smaller, in general, after the two-year interval than were those for the group whose testings were but one year apart. The advantage of a previous acquaintance with Test 3 has largely disappeared over the two-year period, the highest mean scores having been attained by the group tested for the second time after a one-year interval. It is likely that similar results would obtain for the verbal tests were complete data at hand. The effects of the differences of Table 1 will be investigated later.

TABLE 2.—COMPARISONS OF NINTH-GRADE TEST SCORES FOR PUPILS WHO WERE TESTED FOR THE FIRST TIME IN 1946 AND PUPILS WHO HAD BEEN TESTED IN 1944 ALSO

Test	Mean Score		Difference between Means	Standard Error of Difference	Critical Ratio
	Group Tested in 1946 Only	Group Tested in 1944 and 1946			
1 Arithmetic:					
Girls	17.16	16.67	−0.49	0.578	−0.85
Boys	15.27	15.39	.12	.607	.20
All cases	16.23	16.05	− .18	.423	− .43
2 Series Completion:					
Girls	14.53	16.22	1.69	1.048	1.61
Boys	16.49	19.12	2.63	1.211	2.17
All cases	15.50	17.62	2.12	.805	2.63
3 Deduction:					
Girls	29.53	30.40	.87	.868	1.00
Boys	30.62	31.81	1.19	.930	1.28
All cases	30.07	31.08	1.01	0.637	1.59

One of the purposes of the earlier study was to seek any systematic changes which might take place in the factor composition of the tests with increasing maturity. It is here that the short-

comings of a cross-section study are most marked. Not only do pattern weights tend to be sensitive, under some conditions, to minor variations in the correlation coefficients from which they are calculated, but there is also the problem of sampling variations, the full extent of which is not known. In the present instance one source of error is eliminated by the use of scores made by the same individuals on two different testings. In Table 3 are listed the coefficients of the two bi-factors for Group A, who were tested in Grades VIII and IX, and those for Group B, who were tested in Grades VII and IX. Unique factors are not included because the present discussion concerns only the common factors.

The degree to which the several solutions fit the observed intercorrelations is shown in the last three rows of the table. Each of the standard deviations of the final residual correlations is well below the standard error of the mean of the fifteen correlations involved.

There are two aspects of the growth problem which can be considered at this point. First, there has been some question as to whether the factors derived from tests administered at different levels of maturity can legitimately be regarded as the same factors. Two kinds of evidence relating to this question will be presented. If the factors remain the same from year to year, then it should be possible to treat the two independently derived factor patterns for each group as a single pattern with twelve variables instead of six. Except for six specific overlaps, one for each test, this combined pattern should provide an adequate fit to all the sixty-six test intercorrelations. For Group A the mean of the sixty correlations not involving the overlaps is .498, its standard error is .052, and the standard deviation of the final residuals is .036. The corresponding values for Group B are .398, .064, and .060. Thus this type of analysis supports the hypothesis that each of the general and verbal bi-factors obtained at the Grade VII, Grade VIII, and Grade IX levels remains the same factor.

The second kind of evidence that these factors are the same at each level is provided by the correlations of the factor scores which are estimated from the test scores by ordinary multiple-regression methods. There is a regression equation corresponding to each column of Table 3. (See Table 9.) The correlations

of the estimated factor scores are given in Table 4. These coefficients were computed by finding the products of the appropriate standard-score regression equations, summing over the N cases, dividing by N, substituting the observed values of the correlations for all expressions of the form $\Sigma z_a z_b/N$, and making use of the fact that the standard deviations of the regression estimates are the corresponding multiple correlations.

TABLE 3.—BI-FACTOR PATTERNS FOR GROUP A AND GROUP B

Test	Group A (N = 212)				Group B (N = 173)			
	Grade VIII		Grade IX		Grade VII		Grade IX	
	G	V	G	V	G	V	G	V
1	.667	....	.727	....	.622	.....	.596	....
2	.734	....	.850	....	.562	.....	.812	....
3	.636	....	.676	....	.668	.....	.733	....
4	.605	.582	.644	.560	.521	.644	.644	.597
5	.676	.362	.699	.407	.381	.472	.582	.464
6	.555	.695	.569	.703	.312	.795	.535	.591
Total variance	2.52	.95	2.94	.97	1.66	1.27	2.59	.92
Mean correlation	.474		.541		.339		.483	
Standard error of mean correlation	.053		.049		.067		.058	
Standard deviation of residuals	.020		.021		.029		.032	

It should be noted that, although the two factors of any of the four patterns involved are uncorrelated, their estimates are not entirely independent because the estimation is not perfect, as indicated by multiple correlations of less than unity (Table 4). Thus, for example, the general- and verbal-factor estimates from the Grade VIII data for Group A have a correlation of .22, instead of zero. The two estimates of the general factor that were made

for the pupils of Group A have a correlation of .89, which is extremely high in view of the two multiple correlations of .89 and .92. The general-factor estimates for Group B have a correlation of .83. Although this value is lower than that for Group A, the difference may be due largely to the slightly lower multiple correlations and in part to changes that take place during two years of mental growth. Certainly the evidence is clear that a general factor computed from the results of these tests can be given the same interpretation at all three grade levels. The same conclusion may be reached for the verbal factor, although the correlations are lower than those for the general factor. The nature of the bi-factor method of factor analysis is such that the measurement of group factors is somewhat less stable than the measurement of the general factor.

TABLE 4.—INTERCORRELATIONS OF FACTOR ESTIMATES FOR GROUP A AND GROUP B
(Values in diagonals are multiple correlations.)

Group A					Group B				
Grade	Grade VIII		Grade IX		Grade	Grade VII		Grade IX	
	$\overline{G}$	$\overline{V}$	$\overline{G}$	$\overline{V}$		$\overline{G}$	$\overline{V}$	G	$\overline{V}$
Grade VIII:					Grade VII:				
$\overline{G}$	*.89*				$\overline{G}$	*.84*			
$\overline{V}$	.22	*.82*			$\overline{V}$	.18	*.86*		
Grade IX:					Grade IX:				
$\overline{G}$	.89	.25	*.92*		$\overline{G}$	.83	.34	*.91*	
$\overline{V}$	.20	.78	.15	*.85*	$\overline{V}$	.12	.74	.20	*.80*

The second aspect of the problem of changes that might occur with increasing maturity relates to the question of whether or not individuals tend toward increasing use of their 'special abilities' with a corresponding decrease in the use of 'general ability.' Under such a hypothesis, which has been advanced by several

psychologists,[1] the general factor should become relatively less important with increasing maturity. No evidence to support this hypothesis was uncovered in the earlier study. The data of Tables 3 and 4, likewise, fail to supply any supporting evidence. If the hypothesis is true at these levels of maturity, then the relative contribution of the general factor to the total variance should decrease with advancing grade. Obviously, the actual contribution of the general factor increases. Its relative contribution can be assessed by dividing its total variance by that for the verbal factor. The two quotients for Group A are 2.7 and 3.0, and those for Group B are 1.3 and 2.8. These figures not only fail to support the hypothesis—they contradict it. It will not be claimed here, however, that the general factor increases in importance. We prefer to await additional verification of these results, since there is no known method for finding their statistical significance.

The same problem can be considered from another angle by use of the data of Table 4. If it is true that children shift from a generalized method of attack on all mental tasks toward a more specialized procedure, then the general factor for the lower grade might be expected to resemble the verbal factor for the higher grade more closely than the earlier verbal factor would resemble the later general factor. In neither group does this relationship appear. For Group A the correlation of .20 between $\overline{G}_8$ and $\overline{V}_9$ is less than that of .25 between $\overline{V}_8$ and $\overline{G}_9$. The situation is similar for Group B. Moreover, all four of these coefficients differ insignificantly from those already shown to deviate from zero solely by virtue of the fact that the estimates are not perfectly reliable.

A second major line of investigation in the earlier study was based on mean estimates of the factor scores at successive grade levels. Differences found between elementary-school groups and high-school groups could not be evaluated because the

---

[1] Henry E. Garrett, "Differentiable Mental Traits," *Psychological Record* II (June, 1938), 259–98.

Henry E. Garrett, Alice I. Bryan, and Ruth E. Perl, *The Age Factor in Mental Organization.* Archives of Psychology, No. 176. New York: Columbia University, 1935.

Ruth E. Wright, "A Factor Analysis of the Original Stanford-Binet Scale," *Psychometrika,* IV (September, 1939), 209–20.

TABLE 5.—MEAN REGRESSION ESTIMATES OF GENERAL AND VERBAL FACTORS BASED ON NINTH-GRADE DATA FOR EACH OF GROUPS A AND B

Group for Which Mean Regression Estimates Are Presented	Group Whose Regression Equation Is Employed for the Calculation of the Mean Estimates							
	General Factor				Verbal Factor			
	Group A (N = 212)		Group B (N = 173)		Group A (N = 212)		Group B (N = 173)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Group A								
Grade VIII								
100 Girls	44.2	.....	....	.....	45.0	.....	....	.....
112 Boys	44.0	.....	....	.....	46.5	.....	....	.....
All cases	44.1	9.00	....	.....	45.8	9.67	....	.....
Grade IX								
100 Girls	49.9	.....	....	.....	49.8	.....	....	.....
112 Boys	50.1	.....	....	.....	50.2	.....	....	.....
All cases	50.0	10.00	51.0	10.34	50.0	10.00	53.0	9.97
Group B								
Grade VII								
89 Girls	....	.....	41.0	.....	....	.....	45.4	.....
84 Boys	....	.....	43.3	.....	....	.....	45.4	.....
All cases	....	.....	42.1	8.22	....	.....	45.4	9.74
Grade IX								
89 Girls	....	.....	48.7	.....	....	.....	49.9	.....
84 Boys	....	.....	51.4	.....	....	.....	50.1	.....
All cases	49.2	9.36	50.0	10.00	45.9	9.90	50.0	10.00

results of the selection that takes place as the school-leaving age is reached could not be measured and because only a limited number of ninth- and tenth-grade pupils were tested at that time.

For the present data the individuals are the same at both levels. The problem of selection, then, does not appear, but it should be mentioned that the elementary-school results reported here represent only those pupils who do continue in high school and are not necessarily representative of the elementary-school grades in general. The two hundred twelve pupils of Group A constitute eighty-eight per cent of the original eighth-grade group, and the one hundred seventy-three pupils of Group B constitute eighty-four per cent of those who were in the seventh grade in 1944. Not all the missing pupils failed to continue in school—some transferred to other high schools.

In order that comparisons may be made, the multiple-regression equations selected for each group are those which estimate the general and verbal bi-factors from the ninth-grade patterns and correlations. In their original form these equations produce estimates whose mean values are zero and whose standard deviations are equal to the corresponding multiple correlations. Each equation was multiplied by 10/R and the constant, 50, was added in order that the factor mean and standard deviation shall be 50 and 10. The equations were then written in raw-score form for ready substitution. (See Table 10.)

In Table 5 are presented the means and standard deviations of the factor estimates computed from the four ninth-grade regression equations. All the general-factor means in column (2), calculated by substituting appropriate test means in the regression equation based on the ninth-grade data for the two hundred twelve pupils of Group A, are directly comparable. Likewise, the general-factor means in column (4) are based on the ninth-grade data for Group B. The verbal-factor means of columns (6) and (8) are similarly identified. It was possible to avoid the computation of all the individual factor estimates by employing the test intercorrelations in the determination of the standard deviations.[1]

The means for the girls and boys are strikingly similar with the possible exception of the general-factor estimates for Group B,

[1] Swineford, *loc. cit.*, pp. 22–23.

which indicate superiority on the part of the boys. Since this difference is not substantiated by the data for Group A, it is probably a sampling difference only.

The general factor for Group A shows a one-year increase of 5.9 points, or .59 times the ninth-grade standard deviation. The significant differences of Table 1 suggest that part of this increase may be attributed to practice effects on the test scores. Substitution of the mean scores of the pupils tested for the first time in 1945 in the same regression equation yields a mean score of 47.4 for the general factor. It is possible, then, that the actual increase is only 3.3, or one-third of the ninth-grade standard deviation. Making use of the standard deviations of 9.00 and 10.00 in column (3) and the correlation of .881 between the two sets of estimates,[2] the standard error of the difference is found to equal .326. The increase is, therefore, at least ten times its standard error.

The general-factor increase of 7.9 for Group B represents an average annual increase of 3.9. Even after allowance is made for the fact that Group B is a little more homogeneous than Group A in the general factor, although not significantly so, this increase is probably slightly greater than was that for Group A. Unless the discrepancy between the two groups is due entirely to sampling difference, it suggests that the increase from Grades VII to VIII is greater than the increase from Grades VIII to IX—a not unreasonable assumption. Additional data are necessary, however, to establish this relationship.

The groups are similar with respect to the general factor, as indicated by either the two means computed from the Group A equation—50.0 and 49.2 of column (2)—or the two means computed from the Group B equation—51.0 and 50.0 of column (4). Neither difference is as large as its standard error.

The verbal factor for Group A has a one-year increase of 4.2

---

[2] In order to compute this correlation, it is necessary to write the ninth-grade standard-score regression equation in terms of eighth-grade standard scores. This is accomplished by multiplying each regression coefficient by the eighth-grade test standard deviation and dividing by the ninth-grade test standard deviation. The square of this new equation, summed over the two hundred twelve cases and divided by 212, gives the variance of the eighth-grade estimates. The remainder of the calculation parallels that described in connection with Table 4.

(.42 times the ninth-grade standard deviation). Again employing the mean scores of the pupils tested for the first time in 1945, a ninth-grade mean verbal estimate of 50.0 is obtained. The nature of the regression equation is such that the small differences attributed to practice effects tend to cancel one another, and the obtained difference of 4.2 can be accepted with confidence. The correlation between the two sets of estimates is .782, and the standard error of the mean increase is .412.

In the case of Group B, the verbal factor increases by 4.6 points, an average of only 2.3 points each year. This increase is so much less than would be expected that one hesitates to make generalizations on these data alone. All the test results were carefully examined in an attempt to find an explanation. It was discovered that Group B has a ninth-grade mean score on Test 6 which is lower than the Group A ninth-grade mean by four times the standard error of the difference. Test 6 is a vocabulary test. A definite inferiority on the part of Group B in knowledge of word meanings renders these pupils' limited growth in the verbal factor entirely comprehensible. The two groups do differ significantly in the ninth grade, as shown by the difference, $50.0 - 45.9 = 4.1$ from column (6) of Table 5, and the difference, $53.0 - 50.0 = 3.0$ from column (8). Both differences are statistically significant beyond the one per cent level. Furthermore, it is seen from column (6) that the verbal-factor attainment of Group B in the ninth grade (45.9) is almost identical with that of Group A in the eighth grade (45.8). It is unfortunate that the two groups are so dissimilar with respect to the verbal factor that conclusions cannot safely be drawn.

SUMMARY

Scores are analyzed on six tests administered twice to two groups of pupils, of whom two hundred twelve (Group A) were tested when they were in Grades VIII and IX and one hundred seventy-three (Group B) were tested in Grades VII and IX.

Small but statistically significant practice effects are found in four of the tests after the one-year interval and in one of the first three tests after the two-year interval.

A most significant finding is that the factor composition of the six tests does not change materially during the two years repre-

TABLE 6.—TEST MEANS AND STANDARD DEVIATIONS

Test	Group A				Group B			
	Grade VIII		Grade IX		Grade VII		Grade IX	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	17.15	5.019	16.34	5.192	15.01	4.742	16.28	4.858
2	11.46	8.790	17.84	10.357	10.71	8.477	18.60	9.480
3	29.19	6.873	33.50	6.248	28.56	6.950	31.79	7.320
4	29.16	8.053	34.25	7.834	26.51	7.716	33.11	8.454
5	11.18	5.152	14.19	4.934	10.13	4.348	12.97	4.745
6	17.24	8.027	22.47	8.924	13.70	6.360	18.74	8.437

TABLE 7.—TEST INTERCORRELATIONS FOR GROUP A (N = 212)

Test	Grade VIII						Grade IX				
	1	2	3	4	5	6	1	2	3	4	5
Grade VIII											
1											
2	.482										
3	.452	.447									
4	.396	.431	.405								
5	.452	.523	.404	.620							
6	.347	.440	.346	.741	.627						
Grade IX											
1	.822	.474	.369	.419	.493	.387					
2	.584	.730	.559	.524	.552	.479	.626				
3	.503	.470	.561	.437	.513	.412	.477	.582			
4	.450	.446	.343	.845	.598	.655	.468	.574	.413		
5	.480	.520	.339	.637	.769	.609	.535	.543	.494	.678	
6	.332	.414	.261	.757	.631	.823	.399	.485	.399	.760	.684

TABLE 8.—TEST INTERCORRELATIONS FOR GROUP B (N = 173)

Test	Grade VII						Grade IX				
	1	2	3	4	5	6	1	2	3	4	5
Grade VII											
1											
2	.408										
3	.394	.342									
4	.327	.248	.400								
5	.203	.240	.264	.503							
6	.183	.168	.229	.675	.494						
Grade IX											
1	.775	.372	.406	.373	.201	.140					
2	.533	.648	.487	.508	.376	.319	.555				
3	.342	.418	.549	.439	.457	.312	.390	.577			
4	.308	.277	.367	.825	.504	.574	.384	.511	.487		
5	.269	.234	.265	.636	.678	.474	.348	.415	.486	.652	
6	.247	.272	.207	.711	.514	.732	.294	.431	.427	.698	.585

TABLE 9.—REGRESSION COEFFICIENTS—STANDARD FORM

Test	Group A				Group B			
	Grade VIII		Grade IX		Grade VII		Grade IX	
	$\bar{z}_G$	$\bar{z}_V$	$\bar{z}_G$	$\bar{z}_V$	$\bar{z}_G$	$\bar{z}_V$	$\bar{z}_G$	$\bar{z}_V$
1	.250	−.154	.211	−.161	.298	−.118	.141	−.110
2	.343	−.304	.467	−.379	.250	−.085	.438	−.279
3	.231	−.173	.177	−.156	.373	−.224	.299	−.318
4	.127	.312	.075	.327	.202	.331	.140	.536
5	.201	.036	.198	.055	.086	.111	.104	.195
6	.018	.688	−.005	.728	−.049	.603	.019	.391
Multiple correlation	.888	.820	.924	.847	.835	.863	.905	.801

sented by the data. The general factor identified by the tests administered in Grade VII can be interpreted as the same factor as that identified by the same tests administered one or two years later. The same is true of the verbal factor.

TABLE 10.—NINTH-GRADE REGRESSION COEFFICIENTS—RAW-SCORE FORM

Test	Group A		Group B	
	$\overline{G}$	$\overline{V}$	$\overline{G}$	$\overline{V}$
1	.440	−.366	.321	−.283
2	.488	−.432	.511	−.367
3	.307	−.295	.451	−.542
4	.104	.493	.183	.792
5	.434	.132	.242	.513
6	−.006	.963	.025	.579
Constant term	14.23	33.17	11.27	34.94

No evidence is found to support the view that with increasing mental maturity the general factor plays a less important rôle as special abilities are developed. On the contrary, the general ability represented in the present data increases in both its absolute and its relative contribution to the total test variance. Moreover, there appears no shift from general to verbal factor in the correlations of the general-factor scores at the lower grade with the verbal-factor scores at the higher grade.

Mean factor estimates are presented for girls and boys and the total groups, computed in each instance from the ninth-grade data of the corresponding total group. No important sex differences appear. The one-year increase for Group A is at least ten times its standard error for each factor. The increase for Group B in the general factor is such as to suggest that its growth is negatively accelerated from Grades VII to IX, but the data are not conclusive. The two groups are found to differ too widely in verbal ability to permit similar comparisons with respect to the verbal factor.

APPENDIX

The test means, standard deviations, and intercorrelations are given here in order that the relevant data might be accessible. Also included are all the regression coefficients in original form and the ninth-grade regression coefficients in raw-score form which yield ninth-grade means of 50 and standard deviations of 10.

CATTELL'S STUDY OF PERSONALITY

GODFREY THOMSON
University of Edinburgh

It is exceedingly difficult to give a true picture and a fair evaluation of Raymond B. Cattell's *Description and Measurement of Personality,** which almost simultaneously extorts admiration and provokes condemnation. It is a catalog and classification of numerous researches on personality, and also a psychometric analysis into personality factors which Cattell believes to be the 'source traits' explaining the classification. His Herculean labours as a cataloger and classifier are most commendable. As a catalog his book is indispensable. As a classification it may be productive of criticism, but must in all fairness be praised as an honest piece of hard work His excursions into the mathematics of factorial analysis are sometimes misleading, but they are often suggestive. He is fertile of ideas, although too ready to pontificate about them while they are as yet untested and untried.

He often abuses the English language, speaking for example of 'literal measurements' (pp. 99 and 395), when he merely means measurements or, if you like, actual measurements, as he himself calls them lower down on page 395. For the clumsy and ugly terminology he is probably not himself to blame. No doubt many of the terms are the creatures of others. Nouns can be put together in good English, as in 'hat rack' or 'garden gate,' but to do this to excess or with unsuitable words is reprehensible. 'Source trait' and 'surface trait' are rather good and are, I think, his own, but there are too many examples like 'disposition rigidity,' 'frustration tolerance,' 'goals satisfactions,' or 'need blamescape'! The book has been far too hurriedly pushed through the press, with many resulting blemishes. The work is new and attractive; but it is slipshod. The author refers us to appendices—and there are no appendices! He quotes numerical results from his own articles, which cannot be found there. For example on page 499 he gives a table of the loadings of six variables on factors B and C which he says are from his article 58a

* Raymond B. Cattell. *Description and Measurement of Personality.* Yonkers, N. Y.: World Book Co., 1946, pp. 602.

(*Amer. J. Psychol.* 1945, 58, 69–90). But when I turn to Table III on page 88 of that article I find the different values shown (I have reversed the signs, which are consistent but opposite). I expect the book values come from one of his other rotations, and are not merely errors. But he nowhere explains.

Variable	Book		Article	
	B	C	B	C
3	−.44	−.26	−.41	−.24
4	.45	.30	.43	.36
10	.32	.40	.39	.42
17	.34	.38	.31	.38
27	.30	.39	.24	.37
28	.37	.36	.42	.38

His method in making his classified catalog of traits is to begin with Allport and Odbert's list of some forty-five hundred terms and then to throw together those which are practically synonyms. This leaves him with one hundred sixty terms to which he added eleven from other technical accounts. In further reducing these one hundred seventy-one he "leaves the verdict to the correlation coefficient." These coefficients he finds in fourteen different researches, in which varying numbers of subjects were used, from thirty-one in one research, up to three thousand in a research by Ackerson. The main source is a research by himself undertaken specially for this purpose.

He then charts the traits into clusters, a cluster being a group of traits each one of which correlates at least, say, 0.4 with every other. He finds one hundred thirty such clusters, and since these showed considerable overlap with one another he was able to detect fifty 'nuclear' clusters, a nuclear cluster being a set of traits which appear together in several of the clusters, forming their nucleus. These fifty he classifies further into 'personality sectors,' twenty in number (pp 246–268).

It is rather mystifying to find a little later (p. 294) mention of sixty-two clusters, combined into thirty-five nuclear clusters which are to form the subject of his rating experiment and fac-

torial analysis. Presumably these are a selection from those mentioned above. A reference is given to his article in the *J. Abnorm. Soc. Psychol.*, 1943, but I think the reader is entitled to a little explanation in the book itself.

This rating experiment was performed with two hundred eight men, in thirteen groups of sixteen each. Two raters from each group independently rated their fellows (and presumably each other); and if their ratings did not agree sufficiently (0.60) they were further instructed and made new ratings. A rating rank order was thus obtained in each group, converted into scores one to five assuming normal distribution, and separate correlation coefficients calculated for each group of sixteen men. These were transformed into Fisher's z, averaged, and transformed back to r's, which formed the table of five hundred ninety-five correlations to be analysed. Cattell does not, however, give these correlations anywhere! He refers the reader, on page 301, to 'the Appendix, Table 2,' but, presumably by some oversight, there are no appendices in the book.

The details of the centroid analysis are, he next tells us "set out sufficiently elsewhere." (58a.) When the reader looks up reference 58a, in the *Amer. J. Psychol.*, 1945, he finds no such details there, but only a Table III of "Factor Loadings after Rotation for Oblique Simple Structure,' and a Table IV of "Intercorrelations of Factors." The reader has no means whatever of checking these. The Table III has a footnote saying that "the factor-structure is not set out additional to the factor pattern above." It is in any case very unusual to calculate a factor-structure, but one would have liked the structure on the reference-vectors, and the original centroid structure, to be supplied. We are told that this latter contained fourteen factors, and twelve factors were retained after rotation. The point for the cessation of factoring was determined by McNemar's and by Tucker's criteria and in the above-mentioned article Cattell gives in Table II some figures showing how these were applied, but no means are available for checking them. I cannot help fearing that something is wrong somewhere. I have never in my experience known of correlations based on only two hundred eight men giving anything like twelve significant factors. A correlation coefficient with N = 208 is in any case liable to large sampling error. In the present case the actual basic correlations

were each only calculated from sixteen scores, originally only ranks. One would have liked to see what the thirteen r's were from which each final r was averaged via z. Of the thirteen groups, eight were squads of soldiers, and others "business and professional men, artists and artizans," and "Boston Brahmins, Bohemians, and other rarer types." A research claiming to 'establish' the primary traits of personality ought, in psychology as in any science, to lay all the information before the

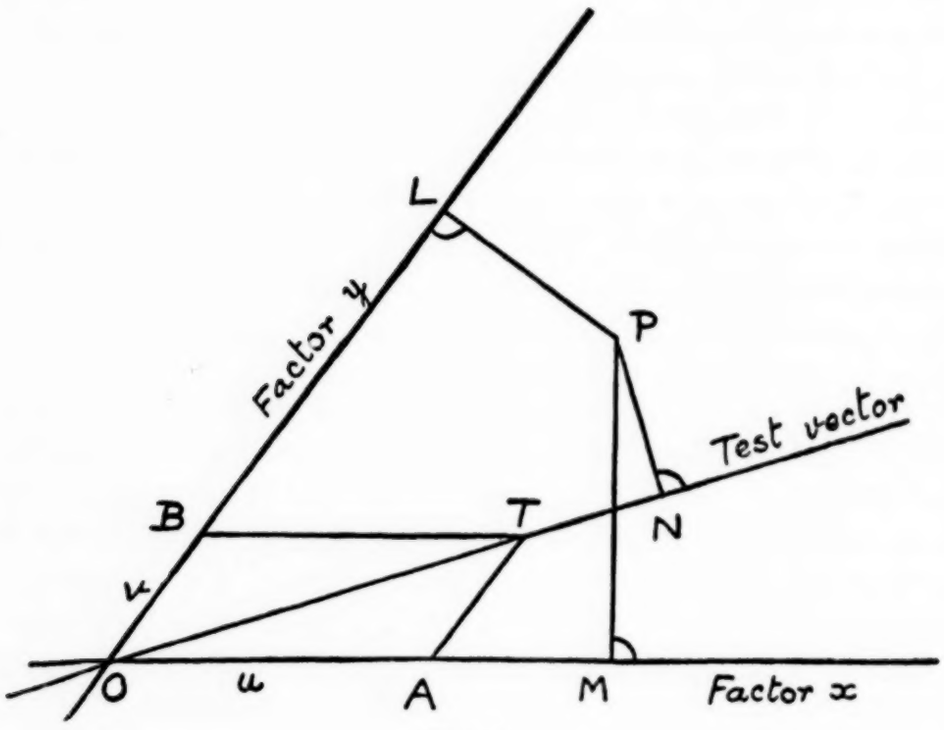

FIG. 1.

reader, and Cattell has not done so. His book is full of ideas and of suggestions and is to be praised for that. But as a scientific report it is very bad indeed.

The above classification and experiment were based on ratings. In chapters 10 and 11 Cattell does something similar, (but without any experiment of his own and without factorial analysis) using self-inventories and objective test measurements. The self-inventories fall into three groups, dealing with neuroticism, etc., with interests, and with attitudes. Again numerous researches have been compared and the traits classified and in some cases compared or even identified with the source traits found in the main work on ratings.

Chapter 12 sums up, and names and describes the twelve 'established' primary traits found earlier, and hints are given of researches now being carried out to help in crystallising these and fixing their relationship. I turn now to some of the mathematical details of the book.

Cattell's page 559 is very confused and gives one no confidence in the mathematical accuracy of his work. He is speaking of the question, "what are the weights to apply to a person's factors to obtain his score in a test?" Now the true answer to this is very simple, and is the same whether the factors are orthogonal or oblique. These weights are proportional to, or with suitable units are equal to, the oblique coördinates of the test point on the factors. In this simplest diagram (Fig. 1) where the test-space is of only two dimensions, if T is a point on the test vector OT such that OT is unity, and x and y are the only two factors, the weights to apply to a person's factors are $u = OA$ and $v = OB$. If P is a person's point, and we drop three perpendiculars, then his score on the test is ON, while he possesses factor x to an amount OM, and factor y to an amount OL. And it can by simple trigonometry be shown that

$$ON = u \times OM + v \times OL,$$

[The interested but non-mathematical reader should try a numerical example by measuring a diagram.]
that is,

Performance in test = u times his amount of factor x
+ v times his amount of factor y

If the factors are orthogonal, then u and v are also the correlations between the test and the factors. But it is not because they are correlations that they appear in such a specification equation. The weights in a specification equation are not in general correlations, though if the factors are orthogonal they happen to coincide with correlations.

I think Cattell means well on his page 559, but he makes some statements which, taken just as he gives them, are erroneous. He has, on page 558, set out the general linear specification equation, which in his notation is

$$PR = S_1T_1 + S_2T_2 + S_3T_3 + \text{etc.}$$

where T_1, T_2 etc. are the amounts of each factor possessed by a given person i; and S_1, S_2 etc. are the weights which have to be applied to them before summing them to obtain this individual's performance PR in a test j. So far so good. But on page 559 he says that this formula represents the correlations of the variable (i.e. the test) with each of the factors. No! The quantities S are not the correlations, unless the factors are orthogonal. They are quite generally the oblique coördinates.

Lower down on page 559 he says "the projections on the factor coördinates are different from the correlations with the factors." They are of course the same when, as is usual, the word projection means projection at right angles. I think he means that the oblique coördinates are different from the correlations. It would be clear if he had said "the 'oblique' projections." Then he talks about reducing "these correlations, rather than the conventional type projections," to zero, a remark which makes me fear that he does not appreciate the difference between the reference vectors and the factors. When we reduce the correlations of tests with a reference vector to zero (that is, reduce to zero the orthogonal projections of the tests on to the reference vector) we at the same time reduce the oblique coördinates on the factor axes to zero. A reference vector is at right angles to a hyperplane. The corresponding factor is the line of intersection of all the other hyperplanes. Cattell's equations at the top of page 560 are in a certain sense correct, but they are the wrong equations. He quotes—

$$\text{Structure matrix} = \text{Pattern matrix} \times \text{matrix of factor intercorrelations}$$

and

$$\text{Pattern matrix} = \text{Structure matrix} \times (\text{matrix of factor intercorrelations})\ -1$$

This is true, but only if the pattern and the structure are on the same coördinates or vectors. Cattell proceeds in his next sentence to say that, therefore, each of the twelve contributions of factors to the variable is itself the sum of twelve products. Not at all. For the above equations are simply not applicable or needed for the purpose, which is to find the pattern 'on the

factors' from the structure 'on the reference vectors.' Thurstone's process arrives at simple structure 'on the reference vectors.' A reference vector is not a factor (unless they are orthogonal). But the pattern on the factors always has zeros in

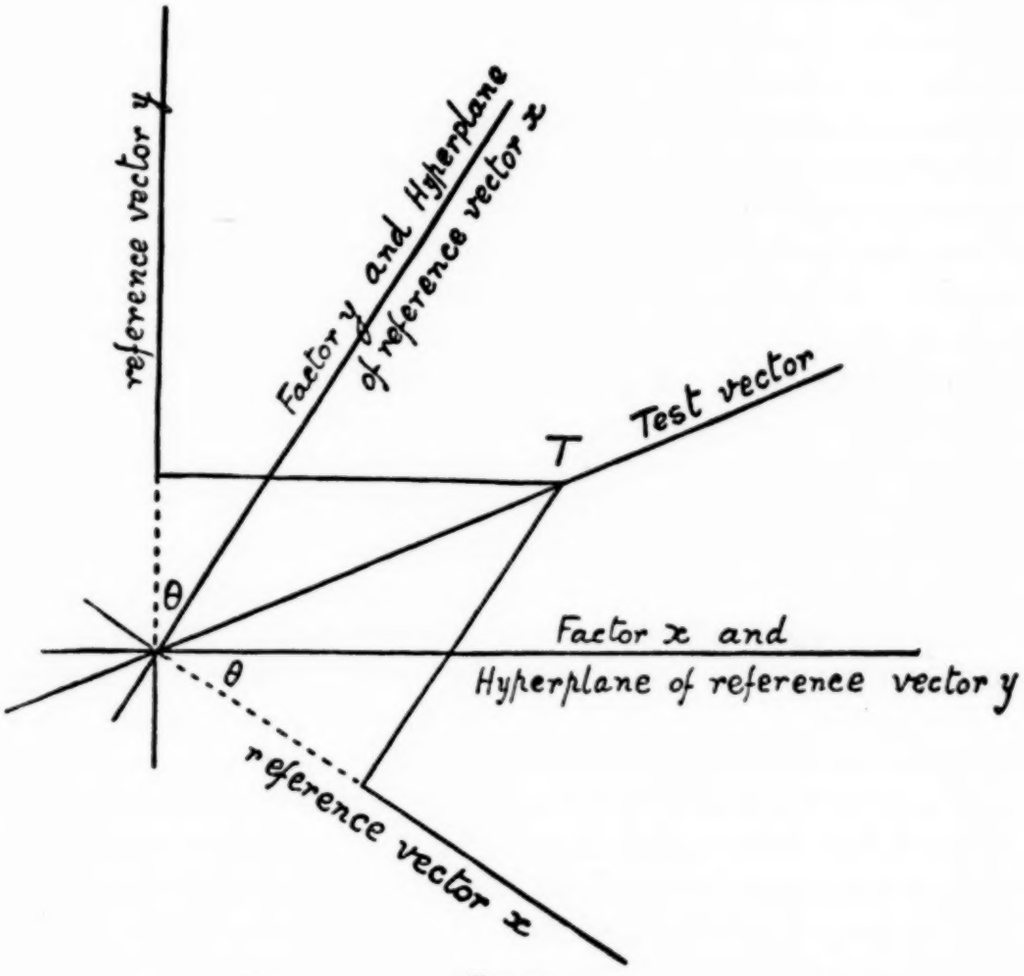

FIG. 2.

the same places as has the structure on the reference vectors. The pattern on the factors (which gives the coefficients in the specification equations, the contributions of each factor to each test) is obtained simply by multiplying each vector column in the reference vector structure by a constant (differing for each column but constant along the column) which is usually not very different from unity. These constants are indeed the reciprocals of the correlations of the reference vectors with the factors, each reference vector with its own factor. And if the factors do not depart very far from orthogonality, each of them will lie close to

its reference vector, so that the cosine of the intervening small angle is not far from unity.

The theory behind all this is considerably illuminated by considering again the extreme case of two dimensions only. In that special case the hyperplane (which is always of one dimension less than the whole) is a line, and moreover the hyperplane of reference vector x is factor y, and vice versa. If T is as before a point on the vector of a test (Fig. 2), the relation between its dotted orthogonal projections on the reference vectors (the structure on the reference vectors) and its oblique coördinates on the primary factors (the pattern on the primary factors) is obvious at sight. The latter are equal to the former divided by the cosine of the angle θ. If one of the above dotted projections is zero for any test, so also will the oblique coördinate on that factor be zero.

As I have already complained, Cattell gives none of the details of his data and calculations which would enable a reader to check his conclusions and diagrams. But it seems probable, from an inspection of his diagram 17 on page 500, that the vectors there labelled 'intelligence' and 'emotionally stable character' are not the factors, but the reference vectors: for he shows the hyperplanes as at right angles to them. Factors are not at right angles to hyperplanes. It is the reference vectors which are. And by the way, what he calls a 'direction cosine' on that diagram is just a cosine, not a direction cosine.

I have felt it a duty to science to call attention to these weaknesses in Cattell's work. But I would like to repeat my admiration of his power of suggesting new ideas. Perhaps the most promising of these is one to which I now turn.

After considering various methods of deciding how to rotate the centroid axes into meaningful positions, Cattell propounds a new principle (previously published by him in *Psychometrika*) of 'parallel proportional profiles,' which, however, his data do not permit him to use in the research described in his present book, so that it is still only an idea. The essence of it is that "if a factor is one which corresponds to a true functional unity, it will be increased or decreased 'as a whole,'" and therefore if the same tests are given under two different sets of circumstance, which favour a certain factor more in one case and less in the other, the loadings of the tests in that factor should all change in the same proportion."

Among 'different circumstances' he mentions different samples

of subjects, differing say in age or sex, and different methods of scoring, or different associated tests in the battery. But he prefers another kind of change of circumstance; namely, a change "from measures of static, inter-individual differences to measures from other sources of differences in the same variables." He instances, among his examples, intercorrelating changes in scores of individuals with time, or intercorrelating differences of scores in twins.

This would apparently work in the following way: First of all, tests a, b , c . . . . . n are given to N individuals, correlated in the usual way, and a centroid analysis made. Then the same tests are given again to these individuals, and the 'changes' in score used to obtain the intercorrelations, which are then subjected to a centroid analysis. Further, the same tests are also given to the twins of the N individuals, and the differences between the twins used as variables.

We now have three centroid analyses, and it is sought to find by rotation positions of the axes such that the profile of loadings of a certain factor is similar in all three matrices. It may even be that the profiles of several factors may be made similar. These positions of the axes then satisfy Cattell's requirement and would be considered by him to indicate true functional unities.

It is not, however, obvious at first sight that the correlated variables are the same in the three cases. In the first case they are the test scores 'simpliciter,' and it is assumed that these are due to some factors or other. In the second case the variables are the 'differences' of score obtained at say Christmas and Easter. The assumption is that if the Easter circumstances favour a factor more than Christmastide did, then this ought to make that factor more influential in each and every test, and produce changes in score accordingly. In practice however I fear that the changes in score will be very unpredictable, and that many other circumstances, as well as the possession of the said factor, will sadly complicate the issue. And something similar will I fear hold for differences in twins' scores, and for many of the other new forms of variable which Cattell so ingeniously portrays in a diagram on page 97. His ingenuity must be admitted. He is productive of ideas, but is too ready to rush into print with them before they have been tested. Nor has he

made, as far as his published work goes, any appreciable progress towards solving the mathematical and arithmetical problems involved in his proportionality idea. His *Psychometrika* article is very elementary, and in the present book he does no more than say in matrix notation what requires to be done. No hint of how to do it. We must hope that this will follow in other publications. But not in such a hurry, and with greater care.

Dr. Cattell will reply to Dr. Thomson's criticism in October or November issue.—Ed.

NOTES ON THE USE OF THE NORMAL DISTRIBUTION IN PSYCHOMETRICS

HOLLIS M. LEVERETT

Graduate School of Education, Harvard University

In the field of mental measurement, there are two distinct applications of the normal curve. The distribution is used (1) as the most probable distribution of chance errors of measurement and (2) as a good approximation of the frequency curve for certain mental traits. In either case, it is customary to make reference to various tables of the normal curve function. Since these tables are not all alike and do not yield the same values, it seems appropriate to note the important differences among them.

THE DISTRIBUTION OF CHANCE ERRORS OF MEASUREMENT

When any measurement is made, it is reasonable to assume that the obtained 'score' does not represent the true value. Repetition of the measure can be expected to yield a somewhat different 'statistic'; that is, measures of the same thing or function are subject to certain variations referred to as errors of sampling or of measurement. Under certain conditions, it is expected that the nature of these errors, that is, the deviations of measures from the true value, assume the distribution form of the normal curve.

Perhaps the simplest and most direct explanation of how the errors of measurement might be expected to take the form of the normal frequency curve is based upon the conditions which yield the symmetrical binomial distribution. This is the familiar coin tossing situation wherein (a) the chance of a positive error (head) is equal to that of a negative error (tail), (b) the quantitative effect of each positive and negative error (each head and tail) is approximately equal, (c) each source of error (coin) acts independently of all others, and (d) the total error in a single measure is the algebraic sum of the effects of all contributing factors (coins). If it can be assumed that the number of causes of error (coins) is very large and each cause contributes a small amount, the binomial distribution becomes a normal curve for all practical purposes.[8]

Since certain measures can be considered as subject to error

from very numerous 'causes' in a comparable manner, it is frequently reasonable to assume that the errors of measurement are normally distributed. Accordingly, we may compute the standard error of measurement and consider it as representing the standard deviation of a normal frequency distribution. Under these circumstances, a table of area values for the normal curve may be used as a probability table. This permits us to raise a number of questions concerning the expected occurrence of certain obtained measures under various hypotheses concerning the true value, that is, the value obtained when the error of measurement is zero. For example, given a particular measure, we may ask what the probability is for that measure to be obtained if the true measure were a given value, X.

For this purpose, there are many tables that may be used. The most thorough of these are in all probability the Kelley-Wood Table[6] and Table I of the Kelley Statistical Tables.[5] Although some tables are more complete than others, there is no question concerning the appropriateness of the values obtained. This is not the case for tables used to normalize a body of data.

THE APPROXIMATE DISTRIBUTION OF MENTAL TRAITS

Since there are no units of measurement for intellect, it is not possible to make an experimental determination of the distribution of a mental trait in any population. All numerical units used in mental measurement are arbitrary, and there is no certainty concerning the equality of these units. This condition holds when measuring instruments use item units, sensed difference units, criterion scale units; regardless of the units involved, the data of mental measurement provide us with relative magnitudes only.

Under these circumstances, the distribution of scores yielded by any test must be either accepted as reasonable or revised to a more appropriate form. Ordered categorical, irregularly multimodal, badly skewed data are often subject to adjustment by converting them to the scale and distribution of the normal curve. The basis for this action need not depend upon a firm conviction that the true function is normally distributed; the data may be normalized properly on the consideration that the revised data are more representative of the true function than the original

obtained data.[7] From this viewpoint, it is not necessary to consider the generally reasonable argument that some mental traits have very numerous determiners which would tend to produce a normal distribution function.

Following the early work of Galton[2] and the publication of the Sheppard table of permille deviates,[3] it became customary to normalize a distribution by assigning median deviate values to certain categorical groupings of the basic data. This practice is followed also when tables like those of Kelley are used.

Although it is not likely that serious errors have resulted from the use of the median deviate, the mean deviate may be considered the more appropriate statistic and its use is recommended. The differences between mean and median deviates may be large or small depending upon the categorical grouping of the data treated; in all cases, however, the use of mean deviates will yield a better approximation of all the characteristics of the normal curve than the median deviates. Accordingly, it is generally desirable to normalize data by using a table of mean deviates like that published by Thorndike[9] or the more detailed table by Leverett.[7]

To demonstrate the differences between mean and median deviates, we may consider the values assigned the following hypothetical five category distribution:

N	*p*	Median Deviates	Mean Deviates
25	.05	1.9600	2.0627
50	.10	1.2816	1.3002
150	.30	.5244	.5422
75	.15	−.0627	−.0631
200	.40	−.8416	−.9659

Other sample deviates may be noted, also, to indicate the general differences existing between mean and median deviates of the unit normal curve. For example:

Portion of Curve	Median Deviate	Mean Deviate
.01	2.5758	2.6652
.50	.6745	.7979
.99	.0125	.0269

From these values, it is readily seen that somewhat different results are obtained if a table of median deviates rather than a table of mean deviates is used. As a rule, the mean deviate is larger than the median deviate for any given portion of the normal curve.

In the case of both the mean and median deviates, it should be emphasized that we are treating with values for the normal curve wherein, theoretically, the entire population is represented. Consequently, if we assign the deviates of certain portions of the curve to the corresponding portions of a sample, it is assumed that the sample is sufficiently large to make every portion of it representative of its population counterpart. This assumption is reasonable for large samples but it is not appropriate for small samples. In small samples, neither the mean nor the median deviates for various portions of the unit normal curve should be used. The phenomenon of regression must be considered.

When treating small samples like that encountered frequently in ranked data of two to fifty cases, Fisher and Yates Table XX[1] is preferred since it does make allowance for regression. Although Hull's Table[4] is frequently used for ranked data, the values yielded by this table are not based upon the expected regression in small samples. The table provides scale scores for ranked data of eleven to fifty cases. The values given can be considered as essentially the deviates of various portions of the normal curve transposed to a scale in which the mean is 5 and the standard deviation is 2. Since the curve was restricted to the range plus and minus 2.5 standard deviations and certain adjustments were made, the values in Hull's table are not comparable to those of other tables.

To demonstrate the differences between tables, we may consider the deviates assigned to typical small samples of ranked data. For this purpose, the values taken from Fisher and Yates Table XX[1], Kelley's Table I[5], and Leverett's Table[7] may be used directly since all deviates are those of the unit normal distribution. The scale values of Hull's table are converted to comparable unit normal deviate values.

The larger differences among deviates may be demonstrated by reference to the values that would be assigned by different tables to the first case in samples of two, three, fifteen, and fifty:

	Deviate for first item			
	Sample Size			
	2	3	15	50
Fisher and Yates.....	.56	.85	1.74	2.25
Hull's Method.....	(.65)*	(.95)*	1.75	2.15
Median Deviate.....	.67	.97	1.83	2.33
Mean Deviate.....	.80	1.08	1.94	2.42

* Estimated.

The larger the sample, however, the smaller are the differences among the deviates. Taking the deviates for a sample of ranked data where N = 20, the following values would be assigned:

Rank	Fisher and Yates Table	Hull's Method	Median Deviate	Mean Deviate
1	1.87	1.85	1.96	2.06
2	1.41	1.40	1.44	1.45
3	1.13	1.15	1.15	1.15
4	.92	.95	.93	.94
5	.75	.75	.76	.76
6	.59	.60	.60	.60
7	.45	.45	.45	.45
8	.31	.30	.32	.32
9	.19	.20	.19	.19
10	.06	.05	.06	.06

It is interesting to note that occasionally Hull's method of restricting the distribution yields deviates which correspond closely to those of the Fisher and Yates Table. This is somewhat accidental, however. The deviates given by Fisher and Yates are the proper statistics since they represent the average value of each case assuming a large number of samples of given size are drawn from a normally distributed population. The median and the mean deviates are generally too large for use with small samples.

In normalizing distributions for large samples, therefore, the

tables of median or mean deviates may be used, but the latter are to be recommended since their use results in a better approximation of the normal frequency curve. Neither the mean nor the median deviates are appropriate for small samples of data. If a sample contains fifty or fewer cases, the table of regressed deviates by Fisher and Yates should be used. For samples of moderate size, it may be necessary to utilize tables of normal curve deviates and this may be done without serious error in most instances; it should be recognized, however, that the deviates assigned are probably a little too large.

SUMMARY

1) The normal curve has two distinct applications in psychometrics. It may be used (a) as the most probable distribution of chance errors of measurement, and (b) as a good approximation of the frequency curve for mental traits.

2) When concerned with a probability function, there are numerous tables of area values for the normal curve; it makes little difference which table is used.

3) When data are to be normalized, however, there are various tables which might be used. These tables do not yield identical values. The differences among the tables are discussed briefly, and it is recommended (a) that mean deviates be used to normalize large samples and (b) that the Fisher and Yates Table be used to assign normalized scale values to small samples.

BIBLIOGRAPHY

1) Fisher, Ronald A., and Yates, Frank. *Statistical Tables for Biological, Agricultural, and Medical Research.* Oliver and Boyd, London; 1938.

2) Galton, Francis. *Natural Inheritance.* Macmillan; 1889.

3) Galton, Francis. "Grades and Deviates," including Sheppard, W. F. "Table of Deviates of the Normal Curve." *Biometrika,* 5, (1907), 400–407.

4) Hull, Clark L. *Aptitude Testing.* World Book Co., Yonkers, N. Y.; 1928.

5) Kelley, Truman L. *The Kelley Statistical Tables.* Macmillan, New York; 1938.

6) Kelley, Truman L. *Statistical Method.* Macmillan, New York; 1923.

7) Leverett, Hollis M. "Table of Mean Deviates for Various Portions of the Unit Normal Curve." *Psychometrika*, June, 1947 141–152.

8) Smith, James G., and Duncan, Acheson J. *Sampling Statistics and Applications.* McGraw-Hill Co., New York; 1945.

9) Thorndike, Edward L. *Introduction to the Theory of Mental and Social Measurements.* Teachers College, Columbia University, New York; 1913.

THE ABILITY OF CHILDREN TO READ A PROCESS-DIAGRAM[1]

MORTON MALTER

Research Fellow, Department of Education

The University of Chicago

INTRODUCTION

Diagrams are probably the most complex and abstract of illustrative materials. As such, educators agree, diagrams should not be used for introducing concepts to elementary-school students. McKown and Roberts[2] sum up this viewpoint in the following statement: "Generally speaking, the pupil should have an adequate 'first look' which . . . gives him a proper initial impression of the thing as a whole in its functional relationships or setting. He should see the item—house, locomotive, manufacturing process, chemical reaction, submarine, solar system, natural development, legislative session, or business organization—first as a complete functioning entity either in its original form or in pictorial or model representation. Following the first look the details of organization, operation, construction and relationships can be shown more easily (in the diagram) and they will be more meaningful."

McKown and Roberts' statement is psychologically sound and describes good pedagogy. Realistically, however, it must be admitted that diagrams are sometimes the first and only nonverbal experience that a child will have with some constructional, organizational or relational concept. Real objects, models or motion pictures, which should preceed the use of diagrams, are not available in many classroom situations. Flat pictures, both photographs and realistic drawings, are generally obtainable but their relationship to diagrams is frequently obscure to the child.

Recognition of the fact that concepts are introduced to children through the diagram leaves us with two choices. On the one

---

[1] This is one of a series of studies being conducted in the University of Chicago Center for the Study of Audio-visual Instructional Materials. The investigation was subsidized in part by a grant from Britannica Jr. through its Advisory Committee.

[2] McKown, H. C. and A. B. Roberts, *Audio-visual Aids to Instruction*, New York; McGraw-Hill, 1940, p. 91.

hand, we might insist that a diagram not based on a 'concrete' foundation is likely to be unintelligible to the unsophisticated reader and let it go at that. On the other hand, we might make some effort to construct textbook diagrams in a manner that will enable at least some children to use them for the starting place in the development of an idea. The latter, the writer believes, is the more practicable, and the following two studies were conducted from that point of view.

STUDY I

Purpose: This study was designed to determine the ability of elementary-school children, grades IV–VIII, to follow the flow of material in a diagram of an industrial process. The writer could not find any studies of this particular ability or of any other aspect of the diagrammatic reading proficiencies of elementary-school children. Information of this kind is basic to any program designed to improve the diagrams utilized in children's books.

Procedure: A number of children's books were examined, and a full-page diagram depicting the flow of wheat through a Flour Mill was selected from one of them. The illustration selected was exactly like the one shown in Fig. 1, but the following were not included: The words 'Start Here and Follow Arrows,' 'Follow Arrows' and 'End Here'; the arrows moving across the elevators; and the numbers 1–4. The illustration was selected for two reasons: (a) it appeared to be a good example of an industrial process diagram that children are likely to encounter; and (b) the idea depicted was new to the children selected for the study.

The illustration was planographed and copies of it were distributed to two hundred twenty-seven children, grades IV–VIII, in two schools. One-hundred-thirty of the children attended School A, a public institution in Wisconsin; ninety-seven of the children attended School B, a private institution in Chicago. The children in School B were superior in respect to intelligence.

An entire class was tested at once, and the following directions were given: "Place the letter S on the place in the flour mill at which the wheat is received. Next, show how the wheat moves from the starting place through the mill by marking its path with a pen or pencil. When you get to the places at which the

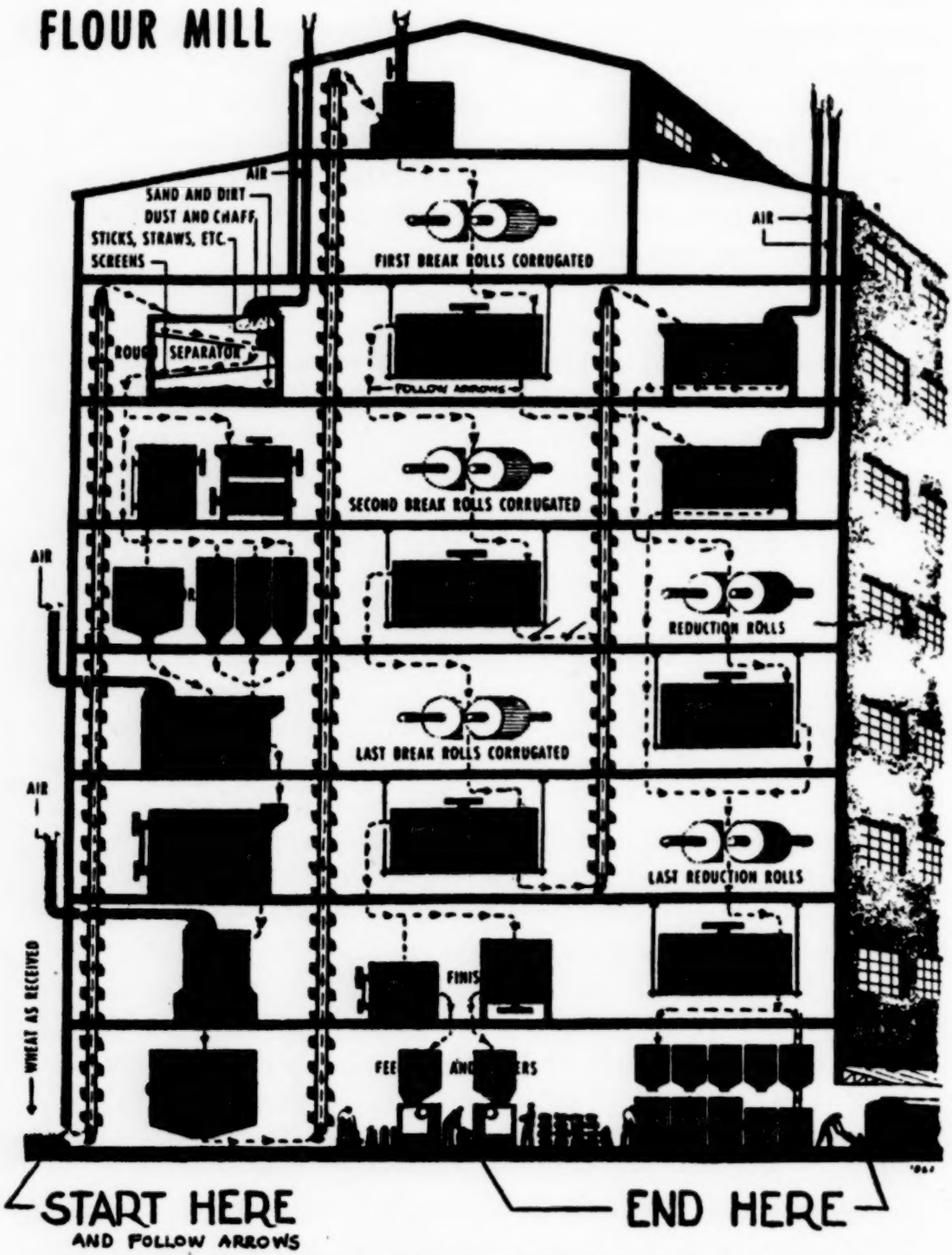

FIG. 1. This illustration was used in Studies I and II. In Study I, however, the following were not present when the illustration was shown to the children: The labels 'Start Here,' 'Follow Arrows' and 'End Here'; and the arrows moving across the elevators. The numbers 1–4 were not present when the illustration was used in both Studies I and II. These numbers indicate the places selected for analysis.

finished products come out, place the letters E there." The students were given all the time that they needed to complete the test. The writer queried the children later, and he believes that they understood the directions.

Results: When the papers were first examined, it appeared that they would have to be scored either totally right or wrong. Many of the children started at the right place but failed to follow the material through to the end. Others started at the wrong place but eventually located the flow of material and moved successfully to the end. Some of the children 'used' the elevators at one place but disregarded them at another. Others failed to follow the flow of material at all, moving at random across the paper. Still others successfully followed the flow of material from start to finish.

TABLE I.—NUMBER OF CHILDREN IN SCHOOL A PERFORMING CORRECTLY ON INDICATED POINT. TABULATED ACCORDING TO GRADES.

Gr.	No. Stu.	Mean CA	Mean IQ	1 Started Correctly	2 Used Elevator	3 First Break Sifter	4 Ended Correctly	5 Totally Correct
IV	34	9.9	102	12	15	8	25	3
V	29	11.8	105	11	18	9	24	5
VI	27	12.0	107	9	12	5	19	3
VII	20	13.2	107	6	5	4	14	2
VIII	20	14.3	113	15	15	8	18	4
Tot.	130	....	...	53	65	34	100	17

It was decided, for the purposes of analysis, to score the papers on each of the following points:

1) Started correctly.

2) 'Used' the elevators at least once.

3) Followed the flow of material in both directions from a place marked 'First Break Sifter.' This was the first place at which the wheat was separated into piles that moved both down and across.

4) Ended correctly. There were several 'end' places, but credit was given for locating any one of them.

5) Totally correct—any paper found correct on each of the above four points. It happens that all papers correct on the first four points were correct in all other respects.

The numbers 1–4 in Fig. 1 indicate the places selected for analysis. Thus, 1 marks the correct starting place; 2, the elevators; 3, the 'First Break Sifter'; and 4, the correct end places.

The total results, tabulated according to school and grade, are shown in Table I and II.

TABLE II.—NUMBER OF CHILDREN IN SCHOOL B PERFORMING CORRECTLY ON INDICATED POINT. TABULATED ACCORDING TO GRADES.

Gr.	No. Stu.	Mean CA	Mean IQ	1 Started Correctly	2 Used Elevator	3 First Break Sifter	4 Ended Correctly	5 Totally Correct
IV	16	9.8	127	6	6	6	14	4
V	21	10.8	137	15	20	13	21	7
VI	23	11.5	135	18	19	14	22	11
VII	15	12.8	126	13	15	8	15	6
VIII	22	13.6	132	22	22	13	22	13
Tot.	97	....	...	74	82	54	94	41

The following results are worthy of note:

1) Only seventeen out of a total of one hundred thirty children in School A were able to follow the flow of material through the mill; forty-one out of a total of ninety-seven children in School B were totally correct.

2) Even the students in the eighth grades of both schools found the test difficult; only four out of twenty in School A and thirteen out of twenty-two in school B were totally correct.

3) Less than half of the children in School A could locate the starting place.

4) Exactly half of the children in School A failed to recognize the function of the elevator.

5) The majority of children in both schools reached the end place; however, many of them reached only one of the end places.

On the whole, the children in School B were more successful on

the test than those in School A. This is not surprising since the children in School B are exceptionally bright and consistently rank a few grades above the norms on most standardized tests.

The following results which cannot be derived from Tables I and II are also worthy of note:

1) Fifty-eight per cent of the School A children who managed to reach the place marked 'First Break Sifter' (3 in Fig. 1) failed to follow the flow of material both down and across.

2) Eighty-six per cent of the children in both schools who failed to locate the starting place started at the upper left-hand corner of the illustration.

Implications of results: Although two hundred twenty-seven children were tested, they were selected from only two schools. Consequently, we do not have a random sample of elementary-school children. Furthermore, the children were tested on only one process diagram. Strictly, then, the findings indicate the ability of particular groups of children to follow the flow of material in a particular diagram.

The children were not shown the verbal material that accompanied the diagram in the book. They might have responded more successfully if they had been allowed to examine this material. However, the writer believes that a full-page process diagram should be as self-contained as possible. It at least should communicate the idea of the flow of material.

Finally, the children were not tested on the total meaning that could be derived from the diagram. They were tested rather on their ability to follow the flow of material in it. The writer does not know to what extent success in following the flow of material is indicative of total understanding. One thing is certain, however: it is impossible for children to understand a process diagram without being able to follow the flow of material through it.

STUDY II

Problem: Study I indicated that the illustration of the flour-milling process was not communicating the idea of the flow of material to many of the children selected for study. On the basis of the findings made in Study I, the writer then altered the illustration by inserting the following: The words 'Start Here and Follow Arrows,' 'Follow Arrows,' and 'End Here'; and the

arrows moving across the elevators. (The numbers 1–4 were not included in this altered illustration.) (Fig. 1.)

Two illustrations were now available: the one used in Study I and the altered illustration. These will be referred to as Illustrations I and II, respectively. This second study was designed to determine if the alterations would help children follow the flow of material through the mill.

Procedure and Results: Planographed copies of Illustration I were presented to thirty-four students selected at random from Grade VI in School C, located in Illinois. The students were given the same directions as were the students in Schools A and B (Study 1). Thus, they were directed to locate the starting place, follow the flow of material through the mill, and locate the end places. The papers were collected at the completion of the test.

Next, planographed copies of Illustration II were presented to the same children, and they were asked to perform as they had on Illustration I. The attention of the children was not directed to the alterations.

The papers were analyzed in the same way that they were in Study I. Total results for Illustrations I and II are shown in Table III.

TABLE III.—NUMBER OF CHILDREN IN GRADE VI OF SCHOOL C PERFORMING CORRECTLY ON INDICATED POINT.

No. Stu.	Mn. IQ	σ IQ
34	94	14.36

	Started Correctly	Used Elevator	First Break Sifter	Ended Correctly	Totally Correct
Illus. I	7	11	7	25	1
Illus. II	23	30	15	33	6

In order to determine if the alterations[3] had significantly affected the mean status of the children tested, the following statistical technique was employed: First, for 'Locating the Starting Place,' each student's paper had been marked right or wrong on 'Locating the Starting Place' in both Illustrations I and II. Thus, two measures were secured for each pupil—one before and the other after the introduction of the alteration.

[3]The alterations are analagous to the 'treatments' employed in method-studies.

The exact hypothesis that we wished to test was that the mean difference would not differ significantly from zero. The value of t was calculated according to the formula $t = \frac{M_O - M_H}{\sqrt{\frac{\Sigma d^2}{n(n-1)}}}$ and equals 5.9. For 33 degrees of freedom, a value of t this large would be found less than one per cent of the time. Thus, we can with reasonable confidence reject the hypothesis stated above.

t's were calculated in a similar fashion for the distributions resulting from the measures secured on (a) Following the Flow of Material across the Elevators, and (b) Following the Flow of Material in Both Directions from the place marked 'First Break Sifter.' Again, the t's secured were sufficiently large to reject the hypothesis that the mean differences were not significant.

t's were not calculated for 'Locating the End Places' and 'Totally Correct.' The pupils did not experience any difficulty in 'Locating the End Places' on Illustration I, and the measure 'Totally Correct' resulted in part from the scores made on the other points selected for analysis.

Many pupils again followed the flow of material in only one direction from the place marked 'First Break Sifter,' thus accounting in part for the relatively small number of totally correct papers for Illustration II. The pupils who made this mistake told the writer that they had not noticed the alteration 'Follow Arrows,' an understandable oversight in view of the fact that this alteration was printed in extremely small letters.

Implication of Results: The results indicated that the alterations significantly increased the performance of the pupils tested. One might argue that it would be foolish to entertain any other hypothesis. For example, it would certainly seem logical to suppose that a child would locate the starting place in an illustration more readily if it were marked 'Start Here,' a term probably meaningful to him.

The writer certainly would agree with this argument. However, the problem we face is essentially one of determining (a) what words and other symbols are most meaningful to the reader, and (b) the proper methods of inserting these symbols. For example, Study II indicated that many children persisted in starting at the upper-lefthand corner in spite of the fact that 'Start Here' was printed in letters about a half-an-inch tall in

the lower lefthand corner. The results also showed that many pupils overlooked the words 'Follow Arrows,' which were extremely small. Obviously, Illustration II can be further altered, and the effects of the alterations can be determined only through testing.

From the standpoint of both time and money, it is impossible to test every process illustration being prepared for publication. This procedure, however, is not necessary. If the factors common to all process diagrams could be isolated and studied, it would eventually result in a number of principles that could be applied in many situations. If anything, Studies I and II indicate that a program of this sort is necessary.

CONCLUSIONS FOR STUDIES I AND II

1) The majority of two hundred twenty-seven pupils tested in Schools A and B were unable to follow the flow of material through a selected illustration of a flour mill.

2) The majority of thirty-four pupils tested in Grade VI of School C were helped in their attempt to follow the flow of material in a selected illustration of a flour mill through the introduction of alterations. These alterations were based on the findings made in Study I.

A STUDY OF THE CONSISTENCY OF INFORMANT RESPONSES TO QUESTIONS IN A QUESTIONNAIRE[1]

JOHN B. GERBERICH

Michigan State College, East Lansing, Mich.

Research workers and other students employing the questionnaire technique for gathering data have become increasingly skeptical concerning the accuracy of the information recorded in questionnaires.

The responses to questionnaires are thought to be somewhat more unreliable than those acquired through other research techniques, because there are so many limitations involved in the use of this method. The questionnaire method of research is open to a number of common objections which in part evolve from the technique and procedure of the individual user. Frequently an attempt is made to solicit information to the extent of several hundred items in a single questionnaire. The use of questions that are leading, that involve elaborate explanation, that are not simple and easily understood, or that elicit biased answers or subjective judgment on the part of the informant is a practice that invalidates the technique. The informant may be careless, indifferent and negligent in answering the questions—a tendency encouraged and promoted by indefinite, involved and non-objective questions. Much of the information asked for is not useable when collected. But even when such known precautions have been observed, the more fundamental question still remains; namely, does any questionnaire give the information requested in all cases, and, if not, in what percentage of instances does it yield accurate data? In short, does the observed inaccuracy spring from faulty use of the questionnaire or does it inhere in the technique? Although the question probably cannot be answered by any single piece of research, specific studies may contribute toward an answer. This paper presents a study of the degree of consistency of informant responses to questions in a questionnaire when both the questionnaire and the technique of administration are controlled as carefully as possible.

[1] Contribution No. 12, Department of Biological Science, Basic College, Michigan State College.

More or less arbitrarily, it was decided that identical questions would be submitted upon three occasions to the same persons. The informants were not told the purpose of the research which was conducted in such a way as to conceal its object. The attempt at concealment proved to be successful. Obviously, five or ten submissions would constitute a more rigorous test of consistency, but for practical reasons such frequency was not feasible. Likewise, there might be a difference of opinion as to whether certain intervals between submissions would be 'long enough' or 'too long.'

In the preparation of the questionnaire a list of sixty questions was compiled from the following: Thurston's "Attitude Questionnaires"; Bell's "Social Adjustment Questionnaires"; and professional application blanks. A conscious attempt was made to secure what would be considered 'good questionnaire questions' on the basis of criteria generally listed in books on research techniques.[2]

Part one of the master questionnaire was composed of factual questions taken from application blanks which are being used by a number of teachers' employment bureaus and superintendents of schools. This group includes eleven questions such as: "What is the date of your birth?"; "What is the population of your home town?"; "What is the date of your last vaccination?"

The second part is based upon social adjustment and contains questions of Home Adjustment, Health and Social Relationships. These questions were taken from Bell's "Adjustment Inventory," widely known and frequently used in many departments of psychology. They were divided into groups containing six questions each; such as, (group one) "Do you ever have a strong desire to run away from home?"; "Is (was) your mother the dominant member of the family?"; and "Do you sometimes feel that you parents are disappointed in you?" (group two) "Do you have many headaches?"; "Are you subject to hay-fever or asthma?"; "Do you take cold rather easily from other people?" (group three) "Do you enjoy social gatherings just to be with other people?"; "Do you find it easy to ask others for help?";

[2] Lundberg, G. A. *Social Research*, New York, Longman's, 1929. Smith H. L. *Educational Research, Principles and Practises*, Educational Publications, Bloomington, Indiana, 1944.

"If you were a guest at a dinner, would you do without something rather than ask to have it passed to you?"

The third part was composed entirely of questions selected from L. L. Thurston's "Attitude Questionnaires": Attitude toward War, Attitude Toward God, A Scale for Measuring Attitude toward the Church, Attitude toward the Negro, and Attitude toward Birth Control. The questions selected from each of these five questionnaires constituted the groups under part three. These five groups were made up of six statements, respectively, except the group on war which is composed of seven items. Examples are: (Group One) "Civil and national differences can be settled without war"; "The disrespect for human life in war is a cause of crime waves." (Group Two) "Although I do not believe in God, I am open-minded about the mysteries of life"; "God is the underlying reality of life." (Group Three) "My church is the primary guiding influence of my life"; "I am neither for nor against the church, but I do not believe that church-going will do anyone any harm." (Group Four) "The Negro should be given the same educational advantages of the white man"; "After the Negro has attained the educational advantages, there would be still an irreconcilable feeling between them." (Group Five) "Scientific methods of birth control are not harmful"; "Birth control information should not be generally available." In this manner it was intended to secure data on the relative consistency of different question types of all parts and groups in the master questionnaire.

The sixty questions constituted the master questionnaire, but it was never submitted to the informants as such. Instead the sixty questions were interspersed among thirty other irrelevant questions in each of the three questionnaires having different titles. The irrelevant questions were used to limit memorization and were chosen to conform with the title of each questionnaire. Each questionnnaire contained the basic experimental questions but was set up as differently as possible from the other two so as not to suggest any connection between them.

The three questionnaires were administered to classes in introductory zoology, hygiene and sociology at the Ohio State University in the summer of 1946. Questionnaires were submitted to a total of one thousand informants in three groups—informant group (a) one hundred at one day intervals; informant group (b)

one hundred at seven-day intervals, and informant group (c) eight hundred at ten-day intervals. Participation was optional with the student and not a classroom requirement. The time of giving the questionnaire was not announced in advance, nor was the problem disclosed until all papers of the third questionnaire were returned. The instructions were printed on the questionnaires and were also given orally so that all the terms and the procedure were clearly defined. Practically no help was given by the administrator after the informants began to answer the questionnaires. All the questions and statements were so selected that the answers would not change during the period of time in which the recording of data took place. The informants were given as much time as they needed to record their responses. Nearly all of them completed their questionnaires in approximately twenty minutes.

Three hundred and forty-three students were either absent one or more of the three days when the questionnaires were administered, refused to participate, or failed to complete the questionnaire. Thus the data were computed for a total of six hundred and fifty-seven informants consisting of seventy-three for informant group (a), eighty-five for informant group (b) and four hundred and ninety-nine for informant group (c).

Each question group was recorded in blocks having three horizontal lines, one for each questionnaire and as many vertical lines as there were statements or questions in each and two additional ones for summaries or totals in each group. The responses for each informant and for each question group, respectively, were placed together so that the consistency could be easily determined. If they were identical for the same question or statement in all three questionnaires, they were considered 'consistent'; but if any of these responses was different the statement was termed 'inconsistent.' No tabulation was made concerning the degree of inconsistency; that is, whether only one of the three replies was inconsistent with the other two, or whether all three were different. The number of consistent and inconsistent questions was tabulated for each group and for each informant. Percentages were determined and the results were recorded.

Informant group (c) was selected, while the problem was still being formulated, to be used for the principle part of the study, because it contained the largest number of informants and was

administered with the longest interval of time. The (c) informant group was composed of two hundred and thirty-two males and two hundred and sixty-seven females. The group was made up of one hundred and seventy-three freshmen, one hundred and forty-two sophomores, one hundred and four juniors, and forty-nine seniors. The range of age varied from seventeen to fifty-six, but only age groups with significant numbers were processed; such as, twenty-nine nineteen years of age, thirty-eight twenty years of age, forty-eight twenty-one years of age, fifty-four twenty-two years of age, fifty-eight twenty-three years of age, fifty-three twenty-four years of age, forty-nine twenty-five years of age, thirty-nine twenty-six years of age, twenty-nine twenty-seven years of age, eighteen twenty-eight years of age and ten at thirty years of age. There were two hundred ninety veterans, of which one hundred six were former officers, and two hundred two non-veterans. Two hundred and forty-three informants had graduated from high school in the upper third of their class, one hundred and sixty-nine in the middle third and thirty-three in the lower third of their high-school class. Informant groups (a) and (b) were approximately proportional in composition to informant group (c).

RESULTS

1) 73.72 per cent of the responses of informant group (c) were consistent. Comparing the present results of consistency with those of Gerberich,[3] 71.86 per cent, and Bains,[4] 76.00 per cent, there does not seem to be a significant difference.

2) Questionnaires were administered to informant groups (a) and (b) at different time intervals than to informant group (c) for the purpose of finding out whether the length of time would affect the consistency of the questionnaire answers. Informant group (a), one-day intervals, showed 91.01 per cent consistency; informant group (b), seven-day intervals, resulted in 76.08 per cent, and informant group (c), ten-day intervals, was 73.72 per cent consistent. The decline of percentage with the increase of

[3] Gerberich, J. B. *A Study of the Consistency of Informant Responses to Questions in a Questionnaire* Unpublished M.A. Thesis, Kent, Ohio: Kent State University Library, 1941.

[4] Bain, Read 1931 "Stability in Questionnaire Responses," *American Journal of Sociology*, XXXVIII, No. 3, pp. 445–453.

the length of time interval would indicate that there is some 'carry over' or memorization, and it appears to follow the normal retention curve.[5]

3) Women exceeded men in consistency by 1.93, a factor hardly statistically significant, although a similar difference was observed by Gerberich.[3] The women exceeded the men in consistency in all types and groups of questions except two groups of the attitudes, war and birth control, and one group of the personal adjustment, the home.

4) Factual questions showed the lowest consistency of the three question types; and through all different groups (age, college rank, veterans or non-veterans, etc.) of questions excepting three which did not vary more than two per cent. These exceptions were scattered unaccountably throughout the attitudinal type.

5) Age seemed not to be a significant factor, although a slight difference (one to ten per cent) did appear; but the variation did not occur in any sequence.

6) The four class groups (freshman, sophomores, juniors and seniors) showed approximately the same consistency, suggesting that the amount of scholastic training and the increasing selectivity operating as students ascend in class ranks apparently do not influence the stability of responses.

7) No informant had all questions consistent, the highest ranking student having six questions of the sixty inconsistent.

8) The veteran-group results show: enlisted personnel, 72.21 per cent consistent; officers, 76.61 per cent consistent; and non-veterans, 73.45 per cent consistent. This would tend to support the contention of the armed services that the more stable men were chosen as officers. From this study evidence points, although not very significantly, to the fact that the armed forces did select the more mentally stable individuals; veterans as a whole (both officers and enlisted men) were 74.41 per cent consistent as compared with non-veterans with a 73.45 per cent consistency.

9) A comparison of the graduating high-school rank (upper, middle and lower third) of informants shows that there exists some evidence that the student who graduates in the upper third of his class is more consistent than the one who graduates in the

[5] Pressey, S. H. and J. E. Janney *Casebook of Research in Educational Psychology*, Harper and Brothers, New York 1937, pp. 394–397.

lower third. It was found that the students who ranked in the upper third of their high-school class were 75.90 per cent consistent, the middle third were 71.28 per cent consistent, and those of the lower third were 67.35 per cent consistent. The differences are not great, but worth noting. Approximately the same proportional difference existed though all types and groups of questions.

IMPLICATIONS FOR QUESTIONNAIRE RESEARCH

In numerous researches as well as in practical data gathering, the collector has traditionally assumed that if he followed sophisticated questionnaire construction techniques he would secure accurate data. There may be those who would question the above statement, insisting instead that the truly sophisticated researcher expects error and allows for it. Nevertheless, in perusing a random sample of ten questionnaire researches published during the last three years not one of them was found to have made either implicit or explicit allowance that his informants may have given incorrect replies.

There is, moreover, a widely current common-sense impression which this and other researches call into doubt; namely, that factual questionnaire information is more reliable than attitudinal or introspective data. The data show that the reverse is true insofar as we may judge accuracy by consistency. The author readily acknowledges that consistency may not guarantee accuracy, but it is certainly obvious that, on the type of factual material which was studied, inconsistency certainly could not denote accuracy. Therefore, the data tend to exaggerate rather than to minimize the incidence of inaccuracy in questionnaire responses of all types, and more especially those pertaining to factual data.

It may be elaborating on the obvious to suggest that additional experiments of this type would assist greatly in standardizing, in some measure, what could be called the margin of probable inconsistency and also very probably the margin of error to be expected in using questionnaire data where large numbers of informants are involved. For the single case or very small group, where margins of error are essentially meaningless, one can but conclude that the questionnaire had best not be trusted for accuracy. There is no reason to assume that questionnaire data are any

less accurate than other forms of communication such as the interview or personal document. This deals with a fundamental imperfection in interpersonal communication. Whether that imperfection is reduced by or accelerated by this device is quite another problem worthy of investigation.

Finally, as has already been briefly mentioned, consistency and accuracy are not the same thing. Here, then, is another research opportunity worth exploring. In broad terms, there seems to be a need for a series of studies along the following lines—A project could easily be planned so that it would yield data pertinent to at least three separate but related problems: (1) consistency of questionnaire responses, (2) accuracy of questionnaire responses, (3) consistency of questionnaire responses with other types of responses; such as, interviews, autobiographical or other undirected forms of communication. Pending such investigations, it is not amiss to suggest that a more cautious acceptance of questionnaire data seems to be indicated.

A COMPARATIVE STUDY OF THE ACADEMIC ABILITY AND ACHIEVEMENT OF TWO GROUPS OF COLLEGE STUDENTS

ARTHUR L. ASSUM

Counseling Center, University of Chicago

AND

SIDNEY J. LEVY

University of Chicago

INTRODUCTION

How personal adjustment is related to scholastic aptitude and achievement is an important problem of concern to those closely associated with college students. It is frequently suggested that the typical maladjusted student is above his more fortunate fellow students intellectually, but that he may or may not fall below the norm in achievement. Some believe the adjustment is made more difficult by lack of scholastic aptitude. It was the desire to gain a clearer picture of the actual relation between personal adjustment and academic aptitude and achievement that prompted this study. We were interested in three problems:

1) How does a group of adjusted students compare with a group of maladjusted students in regard to scholastic aptitude?

2) How does a group of adjusted students compare with a group of maladjusted students in regard to scholastic achievement?

3) What is the correlation of scholastic aptitude and achievement within these two groups?

PROCEDURE

During the year between October 1, 1945, and October 1, 1946, a large number of University of Chicago students used the facilities of the Counseling Center at the University for help in the solution of various kinds of personal problems. That these students came to the Center can be taken as evidence that some degree of maladjustment was present. From this group of students were selected those who were enrolled in the College and had completed counseling at the Center in the period defined above. Of this group, college records were available in the

Examiner's Office for seventy-one students. This group of college students constitutes what will be referred to hereafter as the 'Center Group.'

For purposes of comparison it was necessary to select from the files of the Examiner's Office a group of seventy-one college students who had not been to the Center for counseling prior to or during the same period. These students were equated with the Center Group for: (1) sex, (2) date of entrance into the College, and (3) grade level of entry into the College. Otherwise, selection was at random. This group is called the 'Non-Center Group.' This procedure gave us two groups of students, the Center Group and the Non-Center Group, the former group presumably having a higher incidence of adjustment difficulties than the latter group.

The data gathered for each individual in the two groups consisted of the following—

Ratings of scholastic aptitude given to entering students:

1) A.C.E., Quantitative (Q) Score
2) A.C.E., Linguistic (L) Score
3) A.C.E. Total (T) Score
4) College Reading Ability Score
5) College Writing Ability Score

Ratings of scholastic achievement: These scores are the permanent grades a student receives by taking the comprehensive examinations covering the work of a school year in a given course.

College Comprehensive Examination Scores.

The three problems stated earlier may now be restated in terms of the data used:

1) Is there a statistically significant difference between the Non-Center Group and the Center Group in regard to A.C.E. Q, L, and T scores and the College Reading and Writing Ability scores?

2) Is there a statistically significant difference between the Non-Center Group and the Center Group in regard to the average comprehensive examination scores for each student?

3) What is the correlation between the average comprehensive examination score and the A.C.E. T score made by each student in the two groups?

When the data had been gathered we saw that there was a

difference in the total number of comprehensive examinations taken by the two groups. One further question was raised by this fact:

4) Is there a statistically significant difference in the number of comprehensive examinations taken by the Center and Non-Center Groups?

TABLE I.—RESULTS OF STATISTICAL COMPARISONS

	Mean Center Group	Mean Non-Center Group	SD Cent. Gp.	SD Non-Cent. Gp.	SE_m Cent. Gp.	SE_m Non-Cent. Gp.	Critical Ratio
A.C.E. Quantitative Score	20.261	21.204	3.88	3.77	.460	.447	1.47
A.C.E. Linguistic Score	21.655	20.890	3.81	3.90	.452	.463	1.18
A.C.E. Total Score	21.176	21.106	3.92	3.97	.465	.471	.11
College Reading Ability Score	21.951	21.580	3.60	3.58	.432	.423	.62
College Writing Ability Score	21.359	21.542	4.04	3.75	.479	.448	.27
Comprehensive Scores Averaged[1]	20.134	20.585	3.02	2.81	.357	.334	2.91
Comprehensive Scores Taken as Units[2]	20.33	21.18	3.36	3.69	.18	.16	3.54
Number of Comprehensives Taken	6.73	6.32	3.16	2.75	.38	.33	.82

[1] Comprehensive scores for each student were averaged giving one mean score for each student.

[2] Comprehensive scores were averaged for the group rather than for the student, giving the mean for the group.

RESULTS

It will be noted from the table that the critical ratios indicate no statistically significant difference between the means for the two groups on measures of scholastic aptitude; i.e., A.C.E. Q, L, T scores and the College Reading and Writing scores. However, it is evident that significant differences are present between the means for the two groups in regard to academic achievement as measured by comprehensive scores. When the comprehensive scores for each student are averaged or taken as units, the means for the two groups show a difference that is significant at the one per cent level; that is, the chances are less than one in one hundred such a difference would be obtained by chance alone.

The table shows also that there is not a significant difference in the number of comprehensive examinations taken by the two groups. When the coefficient of correlation was calculated between the average comprehensive examination score and the A.C.E. T score for each student, an 'r' value of .40 for the Center Group and an 'r' value of .35 for the Non-Center Group were found.

CONCLUSIONS

It is interesting to note that the low coefficients of correlation between the A.C.E. T score and the average comprehensive examination score for each student indicate that the former score alone has little predictive value for later achievement. This result is lower than that obtained by the Examiner's Office (usually near .5), and may be due to difference in size of sampling. On the basis of these data it may be inferred that the maladjusted group of students is comparable in academic ability to the adjusted group. However, in regard to academic achievement a difference is noted in favor of the adjusted group.

BOOK REVIEWS

MARIAN J. RADKE. *The Relation of Parental Authority to Children's Behavior and Attitudes.* Minneapolis: University of Minnesota Press, 1946, pp. 123.

The primary purpose of this study is to discover the relationship of different styles of parental authority and discipline to behavior patterns and attitudes of preschool children by the application of some of the more recent methods developed in child behavior studies. A secondary purpose of the study is to explore the value of projective data for investigating the lives of young children. Results include the finding that there is a trend towards less severe discipline and greater respect for child personality in this present generation than in the previous generation and that there are definite relationships between certain home patterns of authority and attitudes and behavior in children. Also reported is a decrease in autocratic methods, associated with improved parent-child relationships, and greater responsibility by fathers.

The subjects were forty-three children enrolled in the nursery school and kindergarten of the Minnesota Institute of Child Welfare and their parents. There were nineteen boys and twenty-four girls, ranging from three years, ten months to five years, ten months. They were mostly from urban homes and represented a select social, economic, and educational sampling. The parents were from professional or business and managerial backgrounds. There was a variety of nationality backgrounds. The average education of the mother was high-school graduation; the average education of the father was the first year of college. There were nine Ph.D.'s and three M.D.'s amongst the fathers. The data were obtained through the use of questionnaires, interview reports, ratings, tests, experimental situations, and projective techniques. With parents and teachers, ratings were used. With children, oral questionnaires and projective techniques were used and also some free association about parents, which was not very free. These data made it possible to investigate the value of methods used and also obtain information relevant for understanding parent-child relationships.

A chapter devoted to a review of projective techniques is sufficient for the purpose though not too comprehensive, and the

review of literature on parent-child relationships is large enough to be nearly exhaustive but surprisingly inexhaustive or erroneous. For example, the investigation of children's attitudes to parents by the use of the free association method of one investigator is summarized with the statement that age and sex differences were not analyzed and no attempt made to relate the responses to child behavior or home situations—this in spite of the fact that easily available studies on sex differences and the rôle of economic security in children's attitudes to parents have been published by this investigator. Though there is much said about the improved methodology, shifting from the consideration of specific factors to a consideration of the effects of contexts and backgrounds of emotional atmosphere, the general methods used and interpretations made are neither dominantly configurational nor dynamic. A population wherein twelve of forty-five parents have Ph.D.'s or M.D.'s is said to represent a normal population. Not only is the sampling not as random as the investigator contends but the generalizations, too, have to be more limited than are given—generalizations about the differences in the two generations certainly hold for this particular group, but whether they also hold for even the upper economic level and upper intellectual level is still questionable. They hold for an extremely select group of people who are around a college campus, where there are opportunities to be connected with an institute of child welfare. It would be surprising indeed if that group did not manifest more knowledge and interest in family relations and child behavior than any random group. All the author is really justified in saying on the basis of her sampling of subjects by the methods used is that in the opinion of a very favored group of parents in a position to learn about child welfare in one of the best institutions in America, they have attitudes which are superior to those manifested by their own parents, who did not have such advantages. The improved attitudes in child training are far from being universal. Certainly the divorce rates and increase in delinquency do not tend to verify the conclusions made here that the generation has improved so much, except perhaps in this particular population.

In spite of this lack of adequacy in covering the literature without awareness, and in spite of the limitations of the population selected, the study does include some interesting facts as well as

reports on methods in the study of children that should be of value to psychologists and psychiatrists doing child guidance work as well as to more informed parents and others who do professional work with children.

H. MELTZER
Psychological Service Center
St. Louis, Missouri

BERTHA B. FRIEDMAN. *Foundations of the Measurement of Values: The Methodology of Location and Quantification.* New York: Bureau of Publications, Teachers College, Columbia University, 1946, pp. 227.

This book recounts some original research in value measurement in a portion of Chapter XV. Obviously the presentation of this experimental work was not the chief object of the book. This purpose could have been served very well by an article in one of the psychological journals. Chiefly, the book is a critique of value behavior and its locations, naming (classification of specific values under the proper heads), and quantification.

Underlying theory of value and underlying theory of motivation are examined both from a philosophical and a psychological standpoint. The discussion of theory of emotion is seemingly made necessary because of the close connection between value, as defined, and emotion. Value is used in the predicate sense 'to value,' as in 'A values X.' "To say A values X is to make a prediction, not a statement about A's present behavior." Reference points and units for value scales and the logic of measurement are critically examined as related to various types of verbal scales and value-behavior scales. Relationships between verbal scale scores and value-behavior scale scores are discussed in the light of published experimental results. The treatment of the various types of scales seems quite competent. Some attempt is made to relate value measurement to basic psychological laws.

The experimental work consists of determining the agreement (1) between 'free responses' to thirteen wish-questions and check list responses made to the same questions; (2) between notebook records (kept by subjects) of " . . . how you spend every minute of your time for two weeks . . . ," and how every cent was spent; and (3) between two verbal tests of how subjects would choose to spend time and money.

J. B. STROUD
State University of Iowa

G. MILTON SMITH. *A Simplified Guide to Statistics for Psychology and Education.* Revised and Enlarged. New York: Rinehart and Co., 1946, pp. 109.

A simplified guide to statistics, such as this, will probably have its greatest use as a kind of 'self-help' booklet that students may study independently or use as a reference. The author, in his foreword, suggests that the volume is primarily intended to "clarify and to supplement the material in general and laboratory courses." He believes, however, that it could serve as a framework for a one-semester course in statistics, since it contains many of the minimum essentials of such a course. This edition of the *Guide* has been enlarged by the addition of sections on tests of significance for small samples, biserial correlation, and chi-square. A number of sections have been revised. Several tables have been added; these include Fisher's table of t and of chi-square. One of the tables that has been added is an original one that makes it possible to estimate readily the standard error of a biserial r. With one exception, all of the tables appear within the text rather than at the back of the volume; consequently, reference to these tables demands searching through the volume.

The attempt to give students a 'short-cut' to statistics often results in setting up a number of precepts that the student is to memorize or to refer to at appropriate times. Presumably these precepts will be adequate to enable him to do something in the majority of situations in which the student finds himself in need of a statistical tool. A tool, however, is a part of an experiment, and its selection and use must be recognized as operations that are part of the whole. The question of whether or not intelligence and ability to read are related can never be answered independently of the techniques used for measuring each of these, and of the techniques used for summarizing these measurements. In a short volume such as this, the precepts are emphasized at the expense of the underlying rationale. As the author says, "In many cases the student is spared the theoretical justification and derivation of these tools. An attempt has been made, however, to make these devices seem plausible." The concept of 'sparing the student' and of 'making devices plausible' suggests the tone of the volume.

This point of view leads to a number of statements in the text that are likely to be misleading to students. The exercises at

the end of Chapter VII, for example, ask the student seven questions that cannot be answered. If the student does, faithfully, work out the percentage of the cases falling above and below certain points in these distributions of achievement test scores that are represented only by means and standard deviations, he is likely to come out with a reinforced idea of the magic of the normal curve. It should not be too much to ask that students in elementary statistics be given some understanding that a normal distribution is a special case and that markedly skewed distributions of test scores and of grades are often desirable. The chapter on grading on the curve, however, reinforces the idea that grades should be normally distributed; and one or two remarks at other points in the volume emphasize this.

Finally, the concept of teaching that is implicit in this volume leaves much to be desired. On page 10, for example, the author says regarding an instructor who is 'supposed' to have given an intelligence test to his new class: "He wants to get a rough idea as to whether he will be teaching just another average class, whether he will have to struggle to get his ideas across, or whether he will have to be on his guard against an excessively bright group of students who will be constantly catching him up on careless slips. He also wants to get a rough notion as to whether the class is of fairly uniform ability or whether there will be a large and annoying difference between the worst and the best students, with the result that what he says may pass over the heads of some but may bore the brighter members of the class." It is hoped that students who expect to teach will not take from this volume the impression that individual differences are annoying, and that addressing remarks to classes constitutes teaching.

CHESTER W. HARRIS
Department of Education, University of Chicago

P. F. VALENTINE, Editor. *Twentieth Century Education.* New York: Philosophical Library, 1946, pp. 655.

Education, being a public enterprise, is always sensitive to public opinion and understanding. Educators may not always be satisfied with their efforts in behalf of better school-community relations, but they can never ignore the significance of the relationship. At the present stage in the nation's history the public view of schools is at a particularly delicate stage, for beneath the

spectacular excitement about better teachers' salaries and expanded educational opportunities for veterans are such serious doubts as, "Can education really lead the way to international peace, or shall we pour our national funds into greater armanents?" or "Will our present school system build sturdy physical health and instill patriotism and character in youth, or must the nation turn to compulsory military service?"

Thirty educators have accepted the present postwar public reexamination of contemporary social institutions as a meritorious undertaking, and have prepared a volume to assist those who seek a better understanding of public education. No attempt has been made to prepare a sales or advertisers' plea, but rather the effort has been to provide a competent, comprehensive, and honest picture. In the editor's words, "[If] it appears that there is too much confusion in the thinking of educators, that is something the [p]ublic should know. On the other hand, if the trend of a truly American scheme appears in these chapters, as the editor believes, then let those who join in the leadership of our democracy strive to strengthen it in the cause of a peaceful and better world" (p. 1). An incidental use for the book is as a helpful overview for students of education, but the primary audience which the book is designed to serve is the general public.

The characteristics which might be expected in a layman's treatment of education appear to be (1) sound scholarship, (2) popular appeal, and (3) functional approach. The first of these three characteristics is undoubtedly the most important, for no amount of popular appeal or functional value can compensate for a treatment which is basically faulty in fact or theory. The thirty educators who prepared the various chapters of *Twentieth Century Education* certainly qualify by reputation and by position as scholars, and the strength of the book they have prepared lies in its scholarship. It is thorough in its coverage, honest in its presentation of differences and agreements among professional educators, contemporary in the facts and viewpoints it provides, and designed with a proper regard for the relative significance of issues. The educational scene in the United States is treated in thirty chapters grouped into five parts as follows: Part I, Theory and Philosophy; Part II, Psychology in Education; Part III, Science and Education; Part IV, Education and Society; and Part V, The School and Its Problems. Within each part, several

authors present their individual viewpoints on specialties to provide either contrasting pictures or a mosaic picture of the area. Thus, in Part I contrasting philosophic backgrounds are presented by such treatments as Merritt M. Thompson's "Idealism as Educational Philosophy," Frederick S. Breed's "A Realistic View of Education," and Arthur D. Fearon's "Catholic Theory." In Part II psychology in education is presented as more of a mosaic than a series of contrasts, by such treatments as Charles L. Jacobs' "The Problem of Formal Discipline," John H. Butler's "Evaluation of Learning," and Norman Fenton's "The Guidance Program."

The scholarly presentation of a large number of important issues in education may make this book valuable for some students of education who seek a survey of the nation's educational system. Unfortunately the qualification of sound scholarship is probably not sufficient to enable the book to accomplish its primary purpose of aiding the public to understand schools. In terms of popular appeal, it appears too academic, too encyclopedic, too unattractive, and too expensive to gain the attention of even intellectuals outside the teaching profession. Best-sellers have been notably lengthy in recent years, but it is doubtful whether many laymen will spend $7.50 for a poorly printed, poorly organized, unstylish book. The quotation used earlier in this review provides an illustration of the printing inadequacies. As illustrative defects in the organization, one may note that a chapter on "The Nature of the Learner" is included under Part III, Science and Education, instead of under Psychology in Education; and an excellent chapter on problems and procedures of early childhood education is included under Part IV, Education and Society, instead of with the chapters on elementary education, secondary education, and higher education in Part V. There are more serious deficiencies, also, in the academic style of the book. Until the Great Books are standard reading diet for the nation's citizens, many passages of the text are likely to repel all but the most persevering readers.

Even a pedantic format and style might be overlooked by an ardent public if citizens could immediately perceive great functional value from reading the book. They might read regardless of difficulties if, for example, the content showed promise of enlightenment on such issues as education and world peace.

education and compulsory military training, education and the economic welfare of the nation, education and juvenile delinquency, education and stable marriage, or education and the tax structure. A book which treated such problems would be psychologically sound. It would serve a need in the nation, and might well attract widespread attention. It might well embody the same discussions of philosophy, psychology, sociology, and school programs, but the focus of these discussions would have meaning and appeal to citizens. The present volume touches on many of the critical public issues in education, but discussions of these vital issues are buried under unrevealing headings and do not constitute the primary emphasis. The material which is provided may serve to induct some laymen into the mysteries of education, and it may give some students an overview of their profession, but its chief contribution to the cause it sponsors will likely be to expose the difficulties of gaining public attention for education. The spadework of making an initial venture into an important area will be well worth while, however, even if it leads only to other efforts which profit by the initial effort.

DAN H. COOPER

University of Chicago

PUBLICATIONS RECEIVED

GORDON W. ALLPORT AND LEO POSTMAN. *The Psychology of Rumor.* New York: Henry Holt and Co., 1947, pp. 247. $2.60.

ROSE H. ALSCHULER AND LA BERTA WEISS HATTWICK. *Painting and Personality: A Study of Young Children.* Two volumes. Chicago: The University of Chicago Press, 1947, pp. 590. $10.00 per set.

HAROLD H. ANDERSON, JOSEPH F. BREWER, AND MARY FRANCES REED. *Studies of Teachers' Classroom Personalities, III.* Follow-up Studies of the Effects of Dominative and Integrative Contacts on Children's Behavior. Stanford University, Calif.: Stanford University Press, 1947, pp. 156. (paper) $2.00.

WALTER G. BOWERMAN. *Studies in Genius.* New York: Philosophical Library, 1947, pp. 343. $4.75.

Career Opportunities in Aviation: A Handbook of Vocational Information. Washington, D. C.: National Council of Technical Schools (839 17th St, N. W.), 1947, pp. 31. (paper) $0.15.

LUELLA COLE AND JOHN J. B. MORGAN. *Psychology of Childhood and Adolescence.* New York: Rinehart and Co., Inc., 1947, pp. 416. $3.50.

CLAUDE V. COURTER, Chairman. *Schools for a New World.* Twenty-fifth Yearbook, American Association of Schools Administrators. Washington, D. C.: American Association of School Administrators, 1947, pp. 448. $2.50.

FREDERICK B. DAVIS. *Utilizing Human Talent.* Washington D. C.: American Council on Education, 1947, pp. 85. (paper) $1.25.

WILLIAM A. HAMM AND OTHERS. *American History and Economics in the High-school Curriculum.* New York: Board of Education, 1947, pp. 98. (paper)

JULIAN HUXLEY. *UNESCO: Its Purpose and Its Philosophy.* Washington, D. C.: Public Affairs Press (2153 Florida Ave), 1947, pp. 62. (paper) $1.00.

E. ERNEST JOHNSON, Chairman. *The Relation of Religion to Public Education: The Basic Principles.* Washington, D. C.: American Council on Education, 1947, pp. 54. (paper)

FRANK E. KARELSEN, JR. *Human Relations: A Challenge to Our*

Public Schools. New York: Frank E. Karelsen, Jr. (230 Park Ave.), 1947, pp. 74. (paper)

H. M. LAFFERTY. *Sense and Nonsense in Education.* New York: The Macmillan Co., 1947, pp. 202.

ROBERT M. LINDER AND ROBERT V. SELIGER, Editors. *Handbook of Correctional Psychology.* New York: Philosophical Library, 1947, pp. 691. $10.00.

JOHN J. LOFTUS AND OTHERS. *Helping Our Young Children to Learn.* Suggestions for Teachers of the First Grade. New York: Board of Education, 1947, pp. 45. (paper)

PAUL MASLOW. *The Analysis and Control of Human Experiences: The Individual Seen through the Rorschach.* Two volumes. (mimoegraphed) Brooklyn, N. Y.: Paul Maslow (16 Court St.), 1946, 1947, pp. 195, 229. (paper)

LOUELLA MILES. *One World in School: A Bibliography.* Montgomery, Ala.: The American Teachers Association, 1946, pp. 58. (paper) $0.35.

J. S. NOFFSINGER, Editor. *The Manual of Standards, and Directory of Private Home Study Schools and Courses.* Washington, D. C.: National Home Study Council, 1947, pp. 48. (paper)

FRANK J. O'BRIEN AND OTHERS. *Food and Nutrition in the Curriculum.* New York: Board of Education, 1947, pp. 68. (paper)

FRANK J. O'BRIEN AND OTHERS. *Procedures in Health Education for Girls in the Secondary Schools of New York City.* New York: Board of Education, 1947, pp. 104. (paper)

HAROLD A. PHELPS. *Contemporary Social Problems.* Third Edition. New York: Prentice-Hall, Inc., 1947, pp. 845. $4.00.

L. L. THURSTONE. *Multiple-Factor Analysis: A Development and Expansion of the Vectors of Mind.* Chicago: The University of Chicago Press, 1947, pp. 535. $7.50.

GERALD F. WEARY. *Democracy's Case against Religious Education on School Time.* Boston: The Beacon Press, 1947, pp. 22. (paper) $0.25.

BEN D. WOOD, Director. 1946 *Fall Testing Program in Independent Schools and Supplementary Studies.* New York Educational Records Bureau, 1947, pp. 58. (paper)